REPORT SERIES IN AEROSOL SCIENCE
N:o 206 (2017)

FROM ELECTRONIC STRUCTURES TO MOLECULAR-LEVEL CLUSTER FORMATION MECHANISMS IN THE ATMOSPHERE

NANNA MYLLYS

Division of Atmospheric Sciences
Department of Physics
Faculty of Science
University of Helsinki
Helsinki, Finland

Academic dissertation

To be presented, with the permission of the Faculty of Science of the University of Helsinki, for public criticism in auditorium A129, A. I. Virtasen aukio 1, on December 1st, 2017, at 12 noon.

Helsinki 2017

Author's Address: Department of Physics
P.O. Box 64
FI-00014 University of Helsinki
nanna.myllys@helsinki.fi

Supervisors: Professor Hanna Vehkamäki, Ph.D.
Department of Physics
University of Helsinki

Docent Theo Kurtén, Ph.D.
Department of Chemistry
University of Helsinki

Jonas Elm, Ph.D.
Department of Chemistry
Aarhus University

Reviewers: Professor Sergey Nizkorodov, Ph.D.
Department of Chemistry
University of California Irvine

Berhane Temelso, Ph.D.
Department of Chemistry
Furman University

Opponent: Josep Anglada, Ph.D.
Departament de Química Biológica i Modelització Molecular
IQAC-CSIC

ISBN 978-952-7091-94-4 (printed version)
ISSN 0784-3496
Helsinki 2017
Unigrafia

ISBN 978-952-7091-95-1 (pdf version)
http://ethesis.helsinki.fi
Helsinki 2017
Helsingin yliopiston verkkojulkaisut

Acknowledgements

In the first place, I thank the people who have helped to get this thesis into its current shape: pre-reviewers Sergey Nizkodorov and Berhane Temelso for careful reviews and useful comments, Hanna Vehkamäki and Theo Kurtén for valuable feedback and proofreads, Jonas Elm for enduring the earliest draft of the thesis and *nonetheless* reading it many times and giving insightful suggestions, and Viivi Hirvonen for correcting grammar and spotting mistakes. European Research Council and Academy of Finland are acknowledged for financial support, the Doctoral Program in Atmospheric Sciences and Nordplus for travel grants, and the CSC-IT Center for Science in Espoo, Finland, for computational resources.

I warmly thank my professor, Hanna Vehkamäki, for giving me the opportunity to work in the group of Computational Aerosol Physics. Thank you for giving me the freedom to create own projects according to my interests and to become an independent researcher. Additionally, I owe my gratitude for my supervisors Theo Kurtén and Jonas Elm. Theo, thanks for suggesting me the area of atmospheric sciences and aiding me especially in the beginning of my studies. In addition to research, you have encouraged me to teach and given the opportunity to lecture. I really respect your knowledge and experience in the field of atmospheric chemistry and in academic career in general. I express my greatest thanks to Jonas. You have been an awesome colleague, office mate, supervisor, friend, and everything else. It has been pleasure to spend a lot of quality time both in and outside work. Thanks for sharing your know-how and passion for computational chemistry with me.

I am very grateful to the Computational Atmospheric Physics group: Olli, Roope, Golnaz, Anna, Monica, Evgeni, Paula, Valtteri, Jacub, Tuomo, and Hanna as well as to the people from the other side of the road: Noora, Siddhart, Heidi, and Theo. I feel thankful for all other co-workers and previous group members. I am happy for all the lunch and coffee breaks we have shared, and of course, countless pints of beer. Thanks for the Division of Atmospheric Sciences for a nice working environment, sähly, and free-time activities. I have really enjoyed my time with you all. I have also had the opportunity to visit many research groups and participate in many conferences and schools all over the world. During these trips I have met a lot of awesome people and I feel grateful to all of you for the wonderful experiences. I especially thank Ilona Riipinen for taking me to be a part of your group. Thanks for your support and inspirational guidance. Thanks Tinja Olenius for your kind assistance in my studies, especially with ACDC. I was always very happy to visit in Stockholm, thanks for that belongs to the whole ACES team.

A special thanks goes to Markku Räsänen for offering me a great opportunity to work in Laboratory of Physical Chemistry already at the beginning of my studies. I was hired as an experimental chemist to study noble gas compounds in solid matrices, and soon I realized that understanding the properties even for an isolated small linear compounds requires knowledge of electronic structure theory. Thus quantum chemistry became an important part of my life. Thanks Masashi Tsuge for helping me with my first computational studies as well as supervising me in the lab. I also thank Lauri Halonen for sharing your enthusiasm for quantum chemistry and spectroscopy. I am grateful for the whole Fysko team for a wonderful time in the past years. Especially Kajsa Roslund and Viivi Hirvonen for organizing with me the *super awesome* Cultural Evenings — and having time after time more insane ideas. I am very thankful to Mikko Muuronen for your friendship as well as assistance in my studies, career planning, and life in general.

My warmest appreciation goes to my lovely friends, who have been my tower of strength outside academia. Thanks Topias Lattu for being my partner in crime for all these years — whatever happens, I can trust that we are on the same side of the bars. I really think highly of you and your inappropriate sense of humour. Thank you Unna Liimatainen for sharing countless memorable moments during past years, I greatly appreciate all support and laughter you give me. I feel very lucky to have found someone, with whom to share similar interests. I also thank Elli, Taru, Viivi, Kajsa, Essi, Niko, Outi, Hamady, football mates, and all other wonderful friends in my life. I am happy to have two lovely kummichildren, Saaga and Luka, who share lots of laughter with me. I am thankful to all the splendid people for making my everyday life better — love you guys!

I am very grateful for my mom, Irma Myllys, for an unquestioned trust, support, and love. You have encouraged me more than I could ever have hoped for. Thanks for staying always on my side! Thank you Tuomo Ponkkonen for your companionship, patience, and interest in my research. Thanks for going through so much with me during these years both in and outside academia. Even when I have been at the darkest corner of quantum chemistry, I have never been alone — you have been there with me.

With love,
Nanna Myllys

Nanna Laura Pauliina Myllys
University of Helsinki, 2017

Abstract

Atmospheric aerosol particles affect the global climate and human health. A large fraction of atmospheric clusters is formed as a result of collisions and favourable interactions between molecules. However, the exact mechanisms and participating compounds are not fully resolved. The cluster formation mechanisms at the molecular-level are essential to understand what kind of effects aerosol particles have on climate change and health-related issues. Currently, aerosol particles provide the largest uncertainties in estimates of the future climate.

In this thesis, potential cluster formation mechanisms between sulfuric acid and oxidized organic molecules with stabilizing compounds are studied using computational methods. Cluster stabilities must be determined accurately in order to provide trustworthy evaporation and formation rates in atmospheric conditions. This leads to the focus of this thesis: to evaluate the accuracy and applicability of different quantum chemical methods, and to find a robust methodology to study atmospheric cluster formation mechanisms and stabilities in the ambient air.

Density functional theory is confirmed to be sufficient to optimize geometries and to calculate vibrational frequencies for molecular clusters. However, for binding energies high-level electronic structure calculations are necessary. The CCSD(T) method is known as the gold standard in quantum chemistry, but it is computationally too demanding for molecular clusters. Therefore, a domain-based local pair natural orbital (DLPNO) approximation is utilized. The DLPNO–CCSD(T) method allows highly accurate calculations for systems comprising more than hundred atoms. The formation energies can be calculated for atmospheric clusters containing up to ten molecules with an approach close to the CCSD(T) accuracy. Large clusters have previously been out of reach with highly accurate quantum chemical methods.

The aim of the theoretical background in this thesis is to present an overview of quantum chemical methods. The introductory part of the thesis can be used as a handbook for problem solving related to molecular-level cluster formation mechanisms. The research presented here contributes significantly to the current knowledge of the participation of organic compounds in the first steps of aerosol particle formation. Additionally, this research suggests that some other mechanisms than clustering, or other chemical compounds are needed to bridge the gap between experimental and theoretical findings. Guidelines for future atmospheric cluster formation studies are given.

Keywords: Atmospheric Cluster Formation, Electronic Structure Theory, Computational Methods, Thermodynamics, Oxidized Organic Compounds, Sulfuric Acid

Contents

Acknowledgements **i**

Abstract **iii**

List of publications **vi**

1 Introduction **1**

1.1 Importance of Atmospheric Aerosol Particles 2

1.2 Motivation and Aims 4

2 Cluster Formation and Growth **5**

2.1 Participating Compounds 5

2.2 Cluster Thermodynamics and Kinetics 10

2.2.1 Cluster Population Dynamics 14

2.3 Theoretical Tools 15

2.3.1 Global Minimum Energy Structure 16

3 Quantum Chemical Methods **18**

3.1 Basis Sets 20

3.2 Wave Function Theory 23

3.2.1 Hartree–Fock 24

3.2.2 Electron Correlation 25

3.2.3 Configuration Interaction 27

3.2.4 Møller–Plesset Perturbation Theory 28

3.2.5 Coupled Cluster 29

3.2.6 Resolution of the Identity 32

3.2.7 Explicit Correlation 33

3.2.8 Local Electron Correlation Approaches 35

3.2.9 Semi-empirical Quantum Chemistry Methods 40

3.3 Density Functional Theory 41

3.3.1 Density Functionals 42

3.4 Thermochemical and Vibrational Analysis 44
3.5 Gibbs Free Formation Energy 49
3.5.1 Towards a Cost-Effective and Robust Approach 49
3.5.2 Review of Computational Approaches used in Atmospheric Clustering Studies 50

4 Results and Discussion **52**
4.1 Computational Methods Assessment 53
4.1.1 Sensitivity and Accuracy of Density Functional Theory 53
4.1.2 Applicability and Robustness of DLPNO–CCSD(T) 57
4.2 Applications 61
4.2.1 Organic Peroxyacid Compounds 61
4.2.2 Multi-Carboxylic Acids 65
4.3 Overview of Papers and the Author's Contribution 72

5 Future Perspectives **74**

References **77**

List of publications

This thesis consists of an introductory review, followed by six research articles. In the introductory part, these papers are cited according to their roman numerals. **Papers I**, **III**, **IV**, and **VI** are reprinted with the permission of American Chemical Society. **Papers II** and **V** are reproduced with the kind permission of Elsevier and The Royal Society of Chemistry, respectively.

I **Nanna Myllys**, Jonas Elm, Roope Halonen, Theo Kurtén, and Hanna Vehkamäki: Coupled cluster evaluation of the stability of atmospheric acid–base clusters with up to 10 molecules. *J. Phys. Chem. A*, 120:621–630, **2016**. DOI: 10.1021/acs.jpca.5b09762.

II **Nanna Myllys**, Jonas Elm, and Theo Kurtén: Density functional theory basis set convergence of sulfuric acid-containing molecular clusters. *Comput. Theor. Chem.*, 1098:1–12, **2016**. DOI: 10.1016/j.comptc.2016.10.015.

III Jonas Elm, **Nanna Myllys**, Noora Hyttinen, and Theo Kurtén: Computational study of the clustering of a cyclohexene autoxidation product $C_6H_8O_7$ with itself and sulfuric acid. *J. Phys. Chem. A*, 119:8414–8421, **2015**. DOI: 10.1021/acs.jpca.5b04040.

IV Jonas Elm, **Nanna Myllys**, Jan-Niclas Luy, Theo Kurtén, and Hanna Vehkamäki: The effect of water and bases on the clustering of a cyclohexene autoxidation product $C_6H_8O_7$ with sulfuric acid. *J. Phys. Chem. A*, 120:2240–2249, **2016**. DOI: 10.1021/acs.jpca.6b00677.

V Jonas Elm, **Nanna Myllys**, Tinja Olenius, Roope Halonen, Theo Kurtén, and Hanna Vehkamäki: Formation of atmospheric molecular clusters consisting of sulfuric acid and $C_8H_{12}O_6$ tricarboxylic acid. *Phys. Chem. Chem. Phys.*, 19:4877–4886, **2017**. DOI: 10.1039/c6cp08127d.

VI **Nanna Myllys**, Tinja Olenius, Theo Kurtén, Hanna Vehkamäki, Ilona Riipinen, and Jonas Elm: Effect of bisulfate, ammonia, and ammonium on the clustering of organic acids and sulfuric acid, *J. Phys. Chem. A*, 121:4812–4824, **2017**. DOI: 10.1021/acs.jpca.7b03981.

Other publications, which are not included in the thesis, are cited according to their roman numerals.

VII Roie Knaanie, Jiri Sebek, Masashi Tsuge, **Nanna Myllys**, Leonid Khriachtchev, Markku Räsänen, Brian Albee, Eric O. Potma, and R. Benny Gerber: Infrared spectrum of toluene: Comparison of anharmonic isolated-molecule calculations and experiments in liquid phase and in a ne matrix. *J. Phys. Chem. A*, 120:3380–3389, **2016**. DOI: 10.1021/acs.jpca.6b01604.

VIII Jonas Elm, **Nanna Myllys**, and Theo Kurtén: Phosphoric acid — a potentially elusive participant in atmospheric new particle formation. *Mol. Phys.*, 115:2168–2179, **2016**. DOI: 10.1080/00268976.2016.1262558.

IX Geoffroy Duporté, Matthieu Riva, Jevgeni Parshintsev, Enna Heikkinen, Luís M. F. Barreira, **Nanna Myllys**, Liine Heikkinen, Kari Hartonen, Markku Kulmala, Mikael Ehn, and Marja-Liisa Riekkola: Chemical characterization of gas- and particle-phase products from the ozonolysis of α-pinene in the presence of dimethylamine. *Environ. Sci. Technol.*, 51:5602–5610, **2017**. DOI: 10.1021/acs.est.6b06231.

X Jonas Elm, **Nanna Myllys**, and Theo Kurtén: What is required for highly oxidized molecules to form clusters with sulfuric acid?, *J. Phys. Chem. A*, 121:4578–4587, **2017**. DOI: 10.1021/acs.jpca.7b03759.

XI Hong-Bin Xie, Jonas Elm, Roope Halonen, **Nanna Myllys**, Theo Kurtén, Markku Kulmala, and Hanna Vehkamäki: Atmospheric fate of monoethanolamine: enhancing new particle formation of sulfuric acid as an important removal process, *Environ. Sci. Technol.*, 51:8422–8431, **2017**. DOI: 10.1021/acs.est.7b02294.

XII Ulrich Krieger, Franziska Siegrist, Claudia Marcolli, Eva Emanuelsson, Freya Gøbel, Merete Bilde, Aleksandra Marsh, Jonathan Reid, Andrew Huisman, Ilona Riipinen, Noora Hyttinen, **Nanna Myllys**, Theo Kurtén, Thomas Bannan, and David Topping: A reference data set for validating vapor pressure measurement techniques: Homologous series of polyethylene glycols, *Atmos. Meas. Tech. Discuss.*, **2017**. DOI: 10.5194/amt-2017-224.

1 Introduction

To be able to see the beauty of the whole world, the surrounding environment, and atmospheric phenomena, one should have an understanding of physics and chemistry, and in particular quantum chemistry. Quantum chemistry provides the physical and mathematical background necessary to describe atmospheric systems on a molecular level. In order to solve complex atmospheric reactions and their influence on the environment, wide interdisciplinary knowledge as well as collaboration between different fields and methodological approaches are required (see Figure 1). The main motivation behind the research in this thesis is to find a high-level computational methodology to describe atmospheric clustering, and use this approach to study molecular-level aerosol particle formation involving oxidized organic compounds. In addition, the aim of the theoretical background presented here is to offer a guide for scientists to gain an overview of quantum chemical methods which may be applied to solve problems related to atmospheric chemical and physical phenomena.

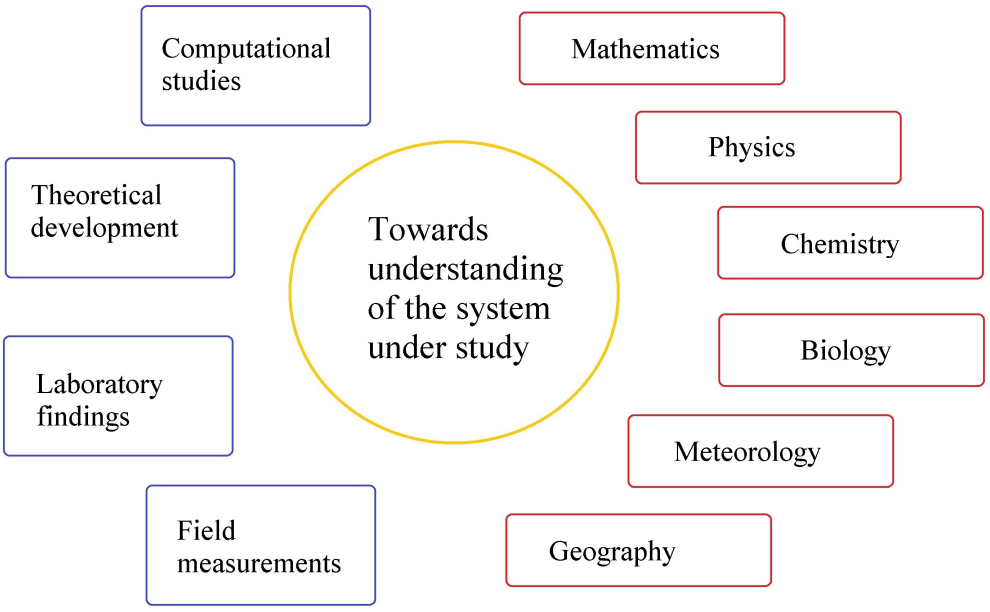

Figure 1: Collaboration between different fields and methodologies enhances understanding of complex phenomena.

1.1 Importance of Atmospheric Aerosol Particles

The air surrounding us can be roughly approximated as a homogeneous mixture of nitrogen and oxygen molecules as well as other non-reactive trace gases. In reality, the Earth's atmosphere contains millions of different chemical compounds from various origins. Some of these are able to form a suspension of liquid, solid, or amorphous particles in the air — an aerosol.[1,2] One cubic centimetre of air contains roughly $3 \cdot 10^{19}$ molecules and from tens up to millions of aerosol particles in the size range from around one nanometre to hundreds micrometers.[3,4] Aerosol particles can be divided into two categories in two different ways: anthropogenic or natural, and primary or secondary.[5] Anthropogenic aerosol particles are produced by human activities, such as industry and traffic, and natural aerosol particles originate from natural sources, such as volcanoes.[6] Primary aerosol particles, such as soot, dust, and sea salt, enter the atmosphere as a particulate matter, whereas secondary aerosol particles are formed in the atmosphere via gas-to-particle conversion.[7,8] It has been estimated that up to half of the particles in Earth's atmosphere are formed via gas-to-particle conversion.[9,10]

Although aerosol particles constitute only a very small fraction of the atmosphere, they influence our planet and humankind in many ways. Aerosol particles directly affect the daily lives of millions of people by degrading air quality.[11,12] Depending on the chemical composition of ultrafine particles, they can harm human health by increasing the risk of cardiovascular and respiratory diseases.[13–15] Particulate matter pollution is linked to an increased risk of lung cancer, irregular heartbeat, heart attacks, and asthma.[15,16] This is a very big concern in, for instance, Chinese mega-cities, where dramatic increase of air pollution events has led to significant enhancement in the incidence of lung cancer.[17] Furthermore, gaseous air pollutants, such as carbon monoxide, ozone, nitrogen oxides, and sulfur dioxide, may seriously affect health.[18] When particulate matter is combined with other air pollutants, the individual effects of each pollutant are amplified. Especially ozone combined with particulate matter has shown to be more harmful than the sum of the individual effects.[19,20] In addition, aerosol particles reduce visibility and, for instance, smog may increase the risk of traffic accidents as well as disturb air and sea traffic. In 2012, the World Health Organization (WHO) assessed air pollution to be responsible for roughly seven million premature deaths in the world per year.[21]

Climate change is one of the most serious threats of the 21st century leading to, for example, melting of the glaciers, warming of the oceans, and the rising of sea levels.[22–24] The consequences will be a challenge for mankind, as some areas could become non-viable for

human life because of increasing drought or flooding.[25–27] The main reason for the observed increase in globally averaged temperatures is the large increase in emissions of anthropogenic greenhouse gases, such as carbon dioxide, methane and nitrous oxides. These gases affect the Earth's radiation budget by absorbing radiation.[28,29] On the other hand, aerosol particles affect the climate via direct and indirect processes, and the effect can be either cooling or warming.[30,31] They can directly influence the radiation balance by either absorbing or reflecting sunlight, and in the case of soot particles by absorbing infra-red radiation emitted by Earth itself.[32] In addition, they influence climate indirectly by acting as cloud condensation nuclei (CCN).[33] Clouds cannot form in the atmosphere without pre-existing CCN (see Figure 2).[34] In the absence of CCN, the formation of cloud droplets from pure water vapour requires an extremely high supersaturation, up to hundreds of percent, which is very unlikely in the atmosphere. Thus cloud droplets are formed on pre-existing water-soluble seeds which are largely originating from secondary aerosol particles.[9]

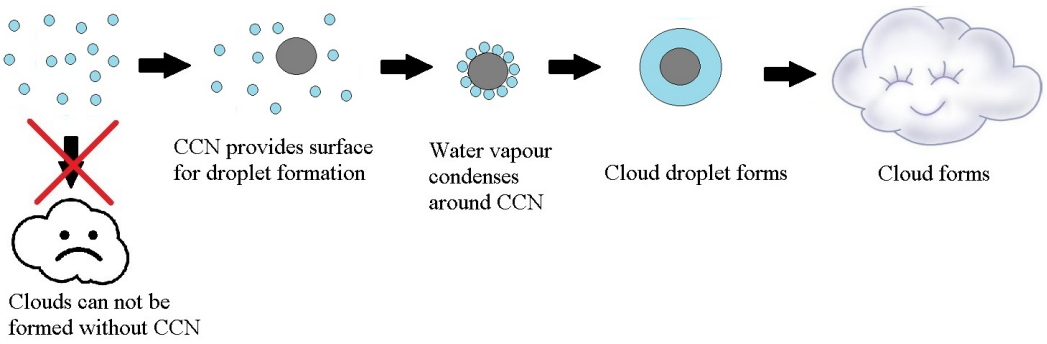

Figure 2: From water vapour via pre-existing cloud condensation nuclei to clouds. Blue circles represent water molecules and grey circles are CCN.

Aerosol particles affect the climate by increasing the number of cloud droplets but reducing their size. This increases the average lifetime and the reflectivity of clouds.[35–37] These effects are assumed to be cooling, however, the quantitative estimates include large uncertainties.[38–40] Aerosol particles are the biggest contributor to the uncertainty of the radiative forcing by aerosol-cloud interactions.[30] Keeping in mind that clouds have an overriding global relevance by regulating Earth's radiative balance, it is clear that an understanding of aerosol particle formation and their properties is needed to solve the worldwide climate issues.

1.2 Motivation and Aims

Although gas-to-particle conversion accounts for a major proportion of atmospheric aerosol particles all over the planet, the fundamental molecular-scale mechanisms and the participation of various compounds remain poorly understood.[41] The formation of aerosol particles from gas-phase molecules is one of the most important and challenging areas of research in atmospheric science. The multitude of proposed cluster formation reactions and participating compounds highlight both the complexity of the phenomena and the large gaps in the current knowledge. Both organic and inorganic compounds have been suggested to have an important role in new-particle formation processes.[42–44] There are instruments capable of counting the number concentrations of sub-3-nanometre particles,[45–47] but resolving the chemical nature of the smallest clusters often require more advanced techniques.[48] In the past years, the development of new high-resolution and high-sensitivity mass spectrometers has increased the knowledge of individual charged clusters at ambient concentrations.[49] In addition, electrically neutral clusters consisting of only a few molecules can be detected using chemical ionization. However, this may alter the composition of the clusters.[48] The details of the initial clustering steps are difficult to probe by experimental means, and therefore, theoretical tools are needed to gain insight into the molecular-scale mechanism.[50]

Quantum chemistry and cluster formation simulations together provide experimentally testable predictions. They are perhaps the most valuable contribution that theoretical chemistry and physics can offer for molecular-level cluster formation studies. The research of this thesis focuses on applying highly accurate quantum chemical tools to studies of atmospheric new-particle formation. Applying state-of-the-art methods is important, as cluster stability depends exponentially on the Gibbs free formation energies, and thus small errors in quantum chemical calculations manifest as large errors in calculated evaporation rates. These challenges have inspired the research of this thesis.

The main objectives of the work of this thesis can be summarized as:
1. Find a robust and cost-effective quantum chemical methodology to study the stability of large atmospheric clusters.
2. Estimate the error in binding energies arising from the domain-based local pair natural orbital approximation.
3. Investigate the molecular interaction of oxidized organic compounds with sulfuric acid.
4. Study the effect of stabilizing compounds which can enhance the cluster formation of oxidized organic compounds and sulfuric acid.

5. Apply state-of-the-art computational methods for molecular clusters and calculate the kinetics of cluster formation in the atmosphere.

2 Cluster Formation and Growth

New-particle formation begins with the collision between individual gas molecules: if the interaction is thermodynamically favourable enough, a molecular cluster can be formed.[41,51] The lifetime of a newborn cluster is extremely short, indicating that the first steps of clustering can be represented as a reversible kinetic process. It should be noted that the equilibrium vapour pressures above a highly curved surface of small clusters are significantly larger than over corresponding bulk liquid due to the Kelvin effect, and therefore the growth of freshly formed clusters is limited and they are often likely to evaporate.[52] However, because a large number of cluster formation and evaporation events occur all the time, some of the clusters can reach critical size, which means that further growth becomes spontaneous.[53] When a particle reaches a size of about 20–100 nm (depending on the chemical composition of the particle), it can act as a cloud condensation nucleus.[39] New-particle formation events which produce a high concentration of aerosol particles are often detected in the lower troposphere in urban, marine, and forested environments as well as in the free troposphere.

2.1 Participating Compounds

Human activities, such as automotive combustion, emit particles into the atmosphere. This influences the concentrations and properties of atmospheric aerosol particles. Anthropogenic processes can also emit condensing or reactive vapours, for example, sulfur dioxide which can oxidise to sulfuric acid in the atmosphere as follows:

$$\mathrm{SO_2 + OH\cdot + M \rightarrow HSO_3\cdot + M} \tag{R1}$$

$$\mathrm{HSO_3\cdot + O_2 + M \rightarrow SO_3 + HO_2\cdot + M} \tag{R2}$$

$$\mathrm{SO_3 + H_2O \rightarrow H_2SO_4.} \tag{R3}$$

Reaction (R3) requires a catalyst such as another water molecule to proceed. Sulfur compounds often occur in relatively high concentrations in fossil fuels, with coal and crude oil deposits commonly containing a few percents of sulfur by weight.[54] The widespread combustion of fossil fuels has greatly influenced atmospheric sulfur emissions, and thus on

a global basis, the anthropogenic emissions are substantially greater than emissions from natural sources.[55]

Sulfuric acid has been observed to be a key component in atmospheric aerosol particle formation, at least over the continental boundary layer.[56] It has been shown that particle formation rates have a strong correlation with the concentration of sulfuric acid.[57–59] However, the atmospheric sulfuric acid concentration is very low (typically 10^{6}–10^{7} cm^{-3}) and it cannot alone produce the observed particle formation events.[60,61] The saturation vapour concentration of sulfuric acid is on the order of 10^{11} cm^{-3}.[62] Water vapour is abundant in the atmosphere (concentration often higher by 8–10 orders of magnitude than that of other condensable vapours)[63,64] and it is likely to participate in many particle formation reactions.[65] Indeed, observations in the free troposphere are well-explained by new-particle formation driven solely by water and sulfuric acid.[66,67] In the boundary layer, however, a binary mixture of water and sulfuric acid has been shown to yield several orders of magnitude smaller formation rates than those observed.[68] It is therefore clear that other stabilizing vapours are needed to produce atmospheric new-particle formation rates at relevant atmospheric concentrations.[62]

Ammonia is the most abundant base in the atmosphere as well as a major constituent of total reactive nitrogen (*i.e.* nitrogen compounds excluding N_2).[69] The major source of ammonia emissions is reported to be agriculture, other sources being industries, vehicular exhausts, vegetation, and oceans.[70–72] In the past, the role of ammonia has been extensively studied, and indeed, it has been shown to enhance the new-particle formation rate in comparison to the binary sulfuric acid–water mechanism.[73,74] Ammonia is able to bind strongly with sulfuric acid by hydrogen-bond formation and proton transfer reactions, and ammonia has been demonstrated to be an important player in atmospheric particle formation.[75,76] Although the ternary new-particle formation of ammonia, sulfuric acid, and water mixture may explain formation events in a specific environment, it is not enough to account for all observed atmospheric formation rates.[77–79]

In addition to ammonia various amines, emitted from both anthropogenic and natural sources, are known to exist in the atmosphere.[80] Approximately 150 amines have been detected in the atmosphere, with alkylamines being the most common ones.[81] The main sources of amine emissions have been identified from human activities such as industrial processes, animal husbandry, fish processing, and landfills.[82,83] Natural sources of amines include marine environments and soils. Amines are basic compounds interacting strongly with sulfuric acid.[84] The gas-phase concentration of ammonia is typically at the level from sub-ppb_V (parts per billion in volume mixing ratio, 10^{10} cm^{-3}) to tens of ppb_V, whereas amines are about

one thousandths of that level.[85,86] Notably the oxidative lifetime of dimethylamine in the atmosphere is of the order of hours, meaning that it mainly has a prominent effect close to sources.[87] Alkyl amines such as dimethylamine are much stronger bases than ammonia and thus bind stronger with sulfuric acid (see **Paper I**).[88,89] A recent study at the Cosmics Leaving OUtdoor Droplets (CLOUD) chamber at CERN confirms that 3 ppt_V of dimethylamine is able to increase cluster formation rates more than a thousand-fold compared to ammonia.[60] This implies that not only the atmospheric abundance but also the basicity and ability to form hydrogen bonds are important when estimating the potential to form stable clusters (see **Paper XI**). It has been demonstrated that diamines can enhance new-particle formation even more effectively than previously studied monoamines.[90] Jen *et al.* used flow tube experiments to show that diamines produce 10 times more particles than dimethylamine and 100 times more than methylamine.[91] In addition, a recent computational study by Elm *et al.* confirms that diamines interact significantly stronger with sulfuric acid than dimethylamine.[92] Putrescine–sulfuric acid clusters are found to be much more ionic than the corresponding dimethylamine clusters, *i.e.*, the proton-transfer ability of diamines is higher than that of monoamines. Therefore, the new-particle formation is significantly more efficient, up to six orders of magnitude compared to the case of sulfuric acid and dimethylamine. Clusters consisting of one or two putrescine compounds and up to five sulfuric acid molecules are found to be stable against evaporation, which is further confirmed by measurements. Elm *et al.* suggests that diamines or other compounds with high basicity might have important role in the initial steps of new-particle formation.[92] More abundant amines with lower basicity can further participate in the particle formation process by attaching to the pre-existing stable clusters. Hence a wide range of amines might be needed to explain the observed new-particle formation events in different environments (see **Paper XI**).

Ions have been suggested to be key players in atmospheric cluster formation.[93,94] Both experimental and theoretical results indicate that ions have a stabilizing role in keeping the condensing species from evaporating easily.[95–97] In laboratory experiments, ions are found to increase nucleation rates by up to ten-fold compared to the neutral cases.[79] The main source of atmospheric ion species is galactic cosmic rays, which frequently ionize gas molecules.[98–100] Also radioactive decay of radon, corona discharge, and lightning are producing ions in the lower troposphere.[101,102] Since ions may be lost by recombination and scavenging by larger particles, the steady-state ion concentration in the lower atmosphere is typically as low as 500–3000 ions per cubic centimetre. Electrically neutral particle formation pathways have been shown to be dominant in the boundary layer in the boreal forest region.[103] In the upper troposphere, the ion concentration is larger, and thus ions are suggested to be important in

new-particle formation.[104] However, there is a some controversy concerning the importance of ions in the lower tropospheric particle formation (see **Paper VI**).[50,105,106]

Bases have been the focus of both experimental and theoretical studies, but the ambient atmosphere also contains various non-basic organic compounds.[107,108] A large quantity of different volatile organic compounds (VOC) are continuously emitted into the atmosphere from anthropogenic and natural sources. The major VOC source is natural emission from vegetation.[109,110] Other reported sources of VOC are oceans, fuel production, biomass burning, and motor vehicles.[111] Terpenes (hydrocarbons with molecular formulas of $(C_5H_8)_n$) constitute a large fraction of biogenic volatile organic compounds (BVOC) in the atmosphere.[112] The most abundant BVOC are isoprene (C_5H_8) and monoterpenes ($C_{10}H_{16}$), which are mainly emitted from plants.[113] It should be noted that when atmospheric temperature increases the emission rates of BVOC are also expected to increase.[114] Figure 3 presents structural formulas of isoprene and several monoterpenes.

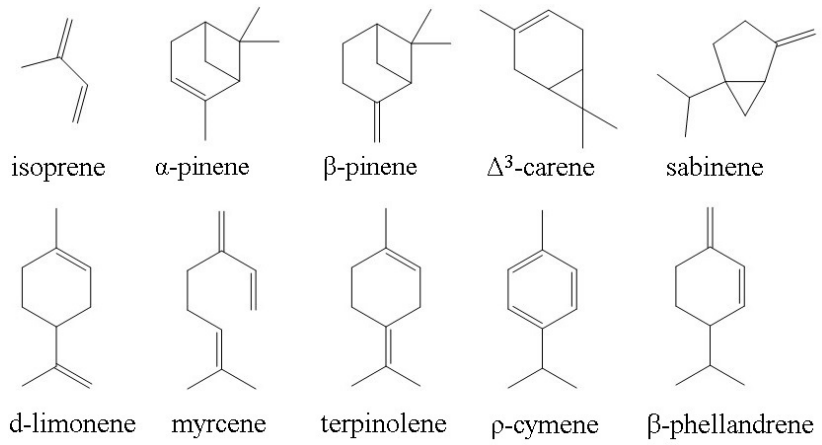

Figure 3: Examples of biogenic volatile organic compounds.

In the atmosphere, BVOC and other volatile organic compounds are oxidised by several oxidants such as the hydroxyl (OH) and the nitrate (NO_3) radicals as well as ozone (O_3).[115] Also the stabilised Criegee intermediates (carbonyl oxides with two charge centres) might affect oxidation of VOC.[116] A large proportion of the aerosol particle growth is believed to originate from low-volatility organic compounds (LVOC).[117–120] However, a large discrepancy between experimentally and theoretically estimated secondary organic aerosol burden from known precursor BVOC implies that there are some missing pieces in the puzzle which could bridge the gap between measurements and models.[117,121,122]

The oxidation products of BVOC cover a broad range of saturation vapour pressures. Only extremely-low-volatility organic compounds (ELVOC) are believed to be capable of participating in the initial steps of clustering.[123–125] Saturation vapour pressures of atmospheric oxidized organic compounds need to be estimated in order to predict partitioning of organic compounds between gas and cluster phases (see **Paper XII**). However, compounds with low vapour pressures at atmospheric temperatures are difficult to probe experimentally. The new high-resolution chemical ionization mass spectrometric instruments are able to detect effectively LVOC and ELVOC (see **Paper IX**).[125,126] In previous studies, they have likely been lost, for example to the chamber walls, before being detected. Recent studies have shown that oxidized organic species participate in the first steps of new-particle formation.[126–128] ELVOC are likely formed through a sequence of unimolecular peroxy radical hydrogen shift reactions and molecular oxygen addition reactions, and eventually a termination reaction leading to closed-shell products.[129] The autoxidation process is believed to produce compounds involving various carbonyl and hydrogen peroxide groups with oxygen-to-carbon ratios above 1 (see **Papers III** and **IV**).[130,131] However, most of the reaction mechanisms and precise molecular structures of oxidized organic compounds are unknown, and thus their role in new-particle formation process is so far poorly understood.[132–134]

As an alternative to autoxidation processes, terpenes can go through several closed cycles of oxidation reactions.[119] After the initial addition reaction, molecular oxygen addition or rearrangements, and termination, the product can be further oxidized by hydrogen abstraction reactions with hydroxyl radicals. From this process several oxidation products of α-pinene, such as pinonaldehyde, pinonic acid, and pinic acid, have been identified.[119,120] Further oxidation of pinonic acid by hydroxyl radicals can yield to the formation of 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA) through complex pathways.[135] It should be mentioned that the consecutive oxidation is a slow process compared to autoxidation, as each generation of products requires another hydroxyl radical for initiating the process. The MBTCA compound represents one of the most promising α-pinene oxidation products to take part in atmospheric clustering process since it contains three carboxylic acid groups (see **Paper X**). **Papers V** and **VI** examine the ability of multi-carboxylic acids to participate in the initial steps of new-particle formation. Figure 4 presents some monoterpene oxidation products identified in the atmosphere.

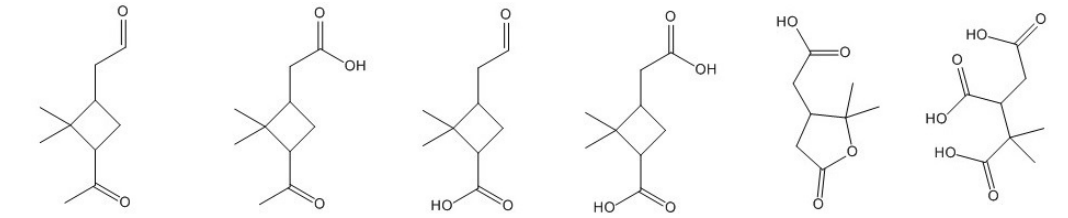

Figure 4: Examples of oxidized organic compounds from α-pinene oxidation.

It should be emphasized that atmospheric new-particle formation is not caused simply by one or two key compounds. The important species behind the phenomenon in different environments differ depending on the vapour concentrations and the relative stabilizing effects. For example, iodine oxide and organic iodine compounds are suggested to contribute to clustering in coastal environments.[136,137] Indeed, a recent study by Sipilä *et al.* found that new-particle formation primarily proceeds by sequential addition of iodic acid on the iodine-rich, coastal atmospheric environment.[138] A numerous possible contributing compounds make the modelling of the clustering process extremely challenging task.[139]

2.2 Cluster Thermodynamics and Kinetics

In atmospheric chemistry and physics, it is important to understand the difference between thermodynamics and kinetics. A common rule in physical chemistry states that kinetics cannot be determined from thermodynamics, indicating that even if the reaction is thermodynamically favourable, the required activation energy can be so high that the reaction is kinetically restricted.[140] At thermodynamic equilibrium, however, it is possible to make a relation between the equilibrium constant and the ratio between forward and backward reaction rate constants.[141,142] Thus by calculating the Gibbs free energies and knowing the other reaction rate constant, the unknown reaction rate constant can be determined.[143] The reaction Gibbs free energy for a cluster C formed from isolated monomers A and B as

$$A + B \rightleftharpoons C \tag{R4}$$

can be obtained from the change of enthalpy ΔH and the change of entropy ΔS at temperature T as

$$\Delta G = \Delta H - T\Delta S, \tag{1}$$

where $\Delta G = G_C - G_A - G_B$. In the clustering process both enthalpy and entropy are decreasing, because the hydrogen-bond formation and proton transferring are exothermic processes, *i.e.*, $\Delta H < 0$ and the flexibility of system decreases when molecules are linked together, *i.e.*, $\Delta S < 0$. This means that for a cluster formation to be thermodynamically feasible at certain temperature, *i.e.* $\Delta G < 0$, the molecular interactions must be strong enough to overcome the entropy penalty (see **Paper X**). It should be noted that when calculating the Gibbs free energies at different temperatures, the enthalpy and entropy are often approximated to be temperature-independent over atmospherically relevant temperature range as in **Papers I**, **V**, **VI**, and **XI**.

Thermodynamic barriers are related to saddle points on the free energy surface, namely *critical clusters*, whereas kinetic barriers are related to first-order saddle points on the potential energy surface, namely *transition states* (see Figure 5).[140] A saddle point is a maximum in one direction of the surface, and a minimum in all other directions.

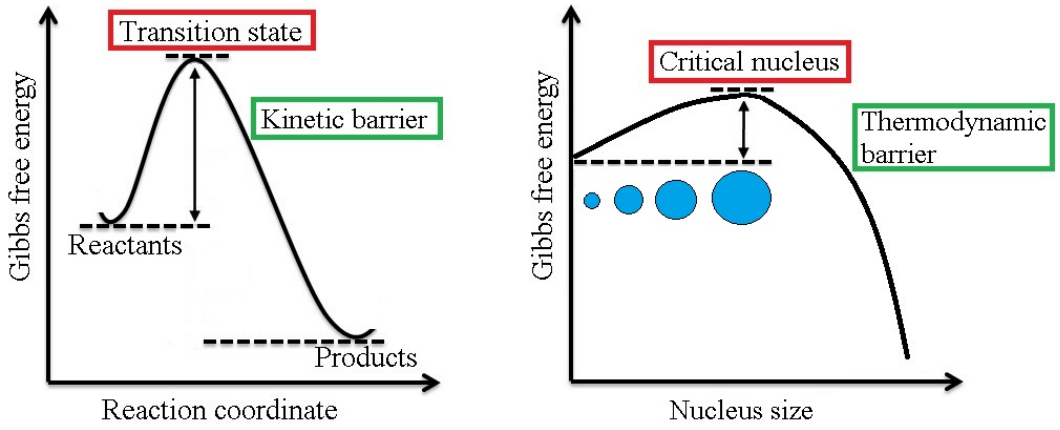

Figure 5: Kinetic (left) and thermodynamic (right) energy barriers.

The separation of kinetic and thermodynamic barriers is not always a straightforward task, since cluster formation may involve real chemical reactions, for instance covalent bonds breaking. These have non-zero kinetic barriers and for this reason ignoring them may lead to a physically wrong description.[144,145] However, for systems in which no chemical reactions occur, it can be assumed that there are no kinetic barriers, and for them the thermodynamic barriers can be used to investigate cluster formation kinetics. Cluster formation via gas-to-particle conversion is generally assumed to proceed via nucleation, *i.e.*, a first order phase transition.[146] This implies that growing clusters must exceed the thermodynamic free energy barrier in order to become stable particles,[41] meaning that the time scale of cluster evaporation

becomes much longer than the rate at which new molecules collide with the cluster. Therefore, cluster growth is favourable after the critical cluster size, which corresponds to the maximum point of the free energy versus cluster size curve.[52] The critical cluster is the smallest cluster for which further growth by collisions with vapour molecules is equally as likely as decay by evaporation.[141,147,148] Hereby, the formation of a critical cluster is a key concept of interest. Alternatively, the formation and growth of clusters can be energetically feasible throughout the cluster size range, meaning that the process is barrierless.[149,150] For instance, the mixture of sulfuric acid and dimethylamine has been suggested to cluster without a free energy barrier (at some theoretical levels and high enough concentrations).[151] **Paper I** shows that high-level quantum chemical calculations predict a barrier for sulfuric acid–dimethylamine nucleation at atmospheric conditions.

A commonly used approach to indirectly deduce the size of the critical cluster is the first nucleation theorem[152–154]

$$\left(\frac{\partial \log J}{\partial \log[A]}\right)_{T,[B]\neq[A]} = n_A^* + \epsilon, \tag{2}$$

where J is the steady-state nucleation rate which depends exponentially on the free energy barrier height, $[A]$ is the gas-phase concentration of compound A, n_A^* is the number of compound A in the critical cluster, and ϵ is a correction term (usually assumed to be negligible). It should be noted that Equation (2) is often derived using classical nucleation theory (CNT), but it can be derived also directly from statistical mechanics.[155] Since the theorem is not limited to CNT, it can be applied independently of the approach used for cluster energies.[141] Several simplifying assumptions have been made in deriving the theorem, and applying it to real atmospheric systems may result in erroneous conclusions concerning the critical cluster size.[153,156] One central assumption behind Equation (2) is that there is one distinct maximum on the free energy versus cluster size curve, which might not be the case for atmospherically relevant systems.[148]

For a cluster to be stable against evaporation at given conditions, it is required that the collision rate is equal to or higher than the evaporation rate.[157,158] According to the law of mass balance, the equilibrium constant K for reaction (R4) can be written as

$$K = \frac{[C]^{\text{eq}}}{[A]^{\text{eq}}[B]^{\text{eq}}} = \frac{k_{\text{B}}T}{p_{\text{ref}}} \exp\left(-\frac{\Delta G}{k_{\text{B}}T}\right), \tag{3}$$

where $[A]^{\text{eq}}$ is the equilibrium concentration of compound A, k_{B} is Boltzmann constant, p_{ref} is the reference pressure at which ΔG is computed, and ΔG is the Gibbs free energy of reaction (R4). At equilibrium, the rate of cluster formation must be equal to cluster destruction, and

thus evaporation $\gamma_{(C\rightarrow A+B)}$ and collision $\beta_{A+B\rightarrow C}$ rates can be calculated as

$$\gamma_{(C\rightarrow A+B)}[C]^{\mathrm{eq}} = \beta_{A+B\rightarrow C}[A]^{\mathrm{eq}}[B]^{\mathrm{eq}}. \tag{4}$$

Here it is assumed that each collision leads to the formation of the product cluster, *i.e.*, the sticking factor, which tells how many of the collisions lead to the cluster formation, is equal to 1. In reality, the colliding species might stick together only when the collision occurs in a specific orientation, and then the collision rate constant as well as the evaporation rate constant are lowered by the sticking factor.[159]

The evaporation rates of clusters can be obtained from Gibbs free energies by assuming detailed balance as in Equation (4):

$$\gamma_{(C\rightarrow A+B)} = \beta_{A+B\rightarrow C}\frac{p_{\mathrm{ref}}}{k_{\mathrm{B}}T}\exp\left(\frac{\Delta G}{k_{\mathrm{B}}T}\right). \tag{5}$$

It should be noted that the reference pressure p_{ref} (usually 1 atm) will cancel out from the evaporation rate, *i.e.*, the evaporation rate is independent of p_{ref}. The detailed balance approach assumes that the evaporation rate of a cluster is independent of ambient conditions other than the temperature and that the cluster settles to the most stable configuration instantly after a collision. However, it is much more likely that structural rearrangements are required to settle to the lowest energy structure, and that the evaporation rate is higher shortly after the cluster is formed compared to the situation after a relaxation time.

The collision coefficients between two electrically neutral clusters can be calculated as hard-sphere collision rates according to kinetic gas theory[160] as

$$\beta_{A+B\rightarrow C} = \left(\frac{3}{4\pi}\right)^{1/6}\left[6k_{\mathrm{B}}T\left(\frac{1}{m_A}+\frac{1}{m_B}\right)\right]^{1/2}\left(V_A^{1/3}+V_B^{1/3}\right)^2, \tag{6}$$

where m_A and V_A are the mass and volume of cluster A, respectively. The volumes are calculated using bulk liquid densities assuming spherical clusters and ideal mixing. Alternatively, the volumes can be calculated from the average radii of the clusters obtained from electronic structure calculations, but the collision rate is not very sensitive to small changes in the cluster volume (see **Paper V**). In the collisions between ions and neutral molecules or clusters, the collision cross section is larger than it would be predicted from the physical dimensions of the colliding systems due to their long-range attraction.[161] This means that particles are interacting with each other already before they collide. Therefore, to get more realistic results, this interaction should be taken into account. This can be done, for example, by applying the approach by Su and Chesnavich, where the masses of the collision partners, and the dipole

moment and polarizability of the neutral collision partner are used to obtain the collision rate.[162,163] A detailed description of that parameterization can be found in **Paper VI** and Ref. 161.

2.2.1 Cluster Population Dynamics

To further investigate the cluster growth into stable aerosol particles and to determine the formation rate, all dynamic processes such as cluster collisions with vapour molecules and each other, fragmentation into smaller clusters, and deposition onto surfaces should be considered. Gibbs free formation energies yield insight into the relative clusters stabilities, but to get information about growth pathways or relative abundances, kinetic effects must be taken into account. The time evolution and behaviour of a population of clusters can be obtained by integrating the time derivatives of cluster concentrations. These birth-death equations (BDEs) include all possible processes where the clusters can be formed or destroyed. The cluster population dynamics can be simulated using the Atmospheric Cluster Dynamics Code (ACDC),[164] which generates the BDEs for a set of clusters, and solves them explicitly by numerical integration using the `ode15s` solver of MATLAB.[165] Usually the stationary steady-state situation is under study when determining aerosol particle formation and then the dynamic simulation can be run until the cluster concentrations are not changing. The BDEs for a cluster C can be written as

$$\begin{aligned}\frac{d\left[C\right]}{dt} = &\sum_{C=A+B}^{A\leq B}\left(\beta_{A+B\rightarrow C}\left[A\right]\left[B\right]-\gamma_{C\rightarrow A+B}\left[C\right]\right)+\\ &\sum_{C=D-A}^{A\leq C}\left(\gamma_{(D)\rightarrow A+C}\left[D\right]-\beta_{A+C\rightarrow D}\left[A\right]\left[C\right]\right)+S_C-L_C\left[C\right]\end{aligned} \tag{7}$$

Here $[C]$ is the concentration of cluster C, $\beta_{A+B\rightarrow C}$ is the collision rate coefficient between A and B, $\gamma_{C\rightarrow A+B}$ is the evaporation rate coefficient of cluster C, S_C is an external source term, and L_C is an external loss term corresponding to coagulation onto pre-existing surfaces. The positive terms are related to formation and negative to the evaporation of cluster C. Figure 6 illustrates the processes affecting the concentration of cluster C. For clarity, the processes are presented for a homomolecular system, but ACDC can be applied to arbitrary multicomponent systems (see **Papers I** and **XI**).

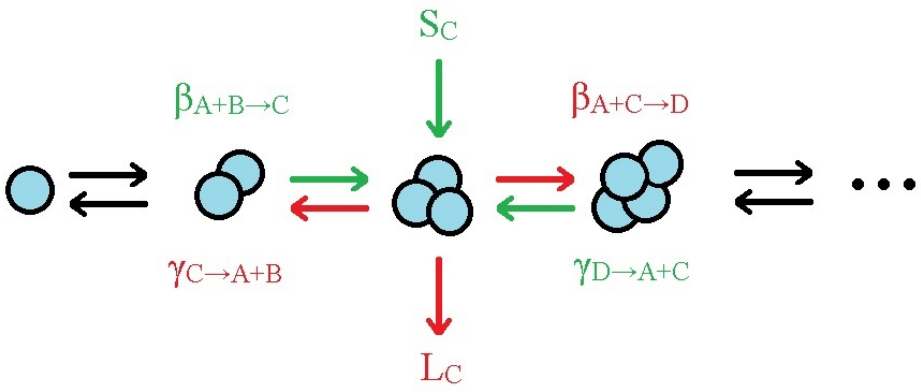

Figure 6: A schematic picture of terms related to formation and destruction of homomolecular cluster C. Green arrows represent positive terms creating cluster C and red arrows are negative terms removing cluster C. Collisions occur from left to right and evaporations from right to left. External source and loss terms are in vertical direction.

2.3 Theoretical Tools

Theoretical approaches to determine the formation free energy change can be classified into three categories depending on the scale at which the interactions between the nucleating compounds are treated: classical liquid drop model,[166,167] force field methods,[168,169] and electronic structure theory.[170] The simplest theoretical framework, classical liquid drop model, treats clusters as spherical bulk-liquid droplets, and requires as input only the liquid density, molecular mass, surface tension, and saturation vapour pressure of the compound.[57] Methods based on force fields — classical density functional theory, molecular dynamics, and Monte Carlo methods — are used to describe the interactions between molecules in terms of their functional groups.[171,172] The basic idea is that a molecular system consists of atoms and their interaction can be described by simple interaction terms without explicitly accounting for quantum mechanical effects.[173] A typical energy expression in molecular mechanics force field methods contains bonded interactions (bond stretching, angle bending, and torsional distortion) and non-bonded interactions (electrostatic and van der Waals).[174]

Macroscopic substance properties or classical force field based methods are usually relatively easy to apply from computational point of view, and thus can be used to study the behaviour of an ideal macroscopic system.[175] Actually, the classical liquid drop model fails for small clusters,[176] and therefore, it should not be applied to draw any quantitative conclusions at the molecular level. To calculate thermodynamic properties of specific atmospheric molecular

clusters, electronic structure calculations are mandatory.

2.3.1 Global Minimum Energy Structure

When using electronic structure theory to obtain thermodynamic properties, common assumptions are that the cluster formation Gibbs free energies can be used as an indicator of the cluster stability, and that the global minimum energy cluster structure dominates the atmospheric cluster distributions, thereby offering an approximative way to describe the properties of a multitude of clusters. The central question becomes how to find the global minimum energy configuration. Unfortunately, there is no simple and universally applicable answer to this. Finding the lowest energy structure for small clusters is usually a fairly straightforward task using just chemical intuition. However, when the cluster contains several molecules, estimating the global minimum energy structure becomes challenging, because the number of local minima on the potential energy surface rises fast with the number of molecules in the cluster.

A commonly-used approach to construct cluster structures is to utilize pair potential molecular dynamics simulations, for example, the simulated annealing technique.[177] **Paper I** uses cluster structures created by simulated annealing.[143] In this three-step model all clusters are simulated at a high temperature starting with random initial configurations to explore the configuration space in depth. After that, molecules are cooled down in order to get real chemical bonded clusters. Finally, clusters are simulated close to absolute zero temperature to reach an energetic minimum structure, which is then used as an initial guess for the density functional theory (DFT)[178] geometry optimization (see Section 3.3 for DFT).[88]

Another common option to obtain the cluster structure is to begin with a large number of different cluster configurations and narrow the pool of clusters incrementally by finding the lowest energy structures with increasingly higher-level computational methods.[179,180] It should be emphasized that different methods yield different potential energy surfaces, meaning that the global minimum energy structure might also be different. **Papers III–VI**, **VIII**, and **XI** apply a semiempirically guided technique where a large number of clusters are created by distributing a molecule or cluster around the target molecule or cluster (see Figure 7). Then all structures are initially optimized using the semiempirical PM6 method[181] (see Section 3.2.9) and single point energies for the converged structures are calculated using DFT with a small basis set. Then the different conformations are identified based on the total energy and dipole moment. The lowest energy conformations are subsequently optimized and vibrational

frequencies are calculated using DFT. Then the conformers are sorted again, and the lowest energy conformers are optimized and frequencies calculated at a higher level of theory. By narrowing the pool of cluster structures systematically, a large portion of the configuration space is sampled, and one should obtain a good guess for the global minimum energy structure.

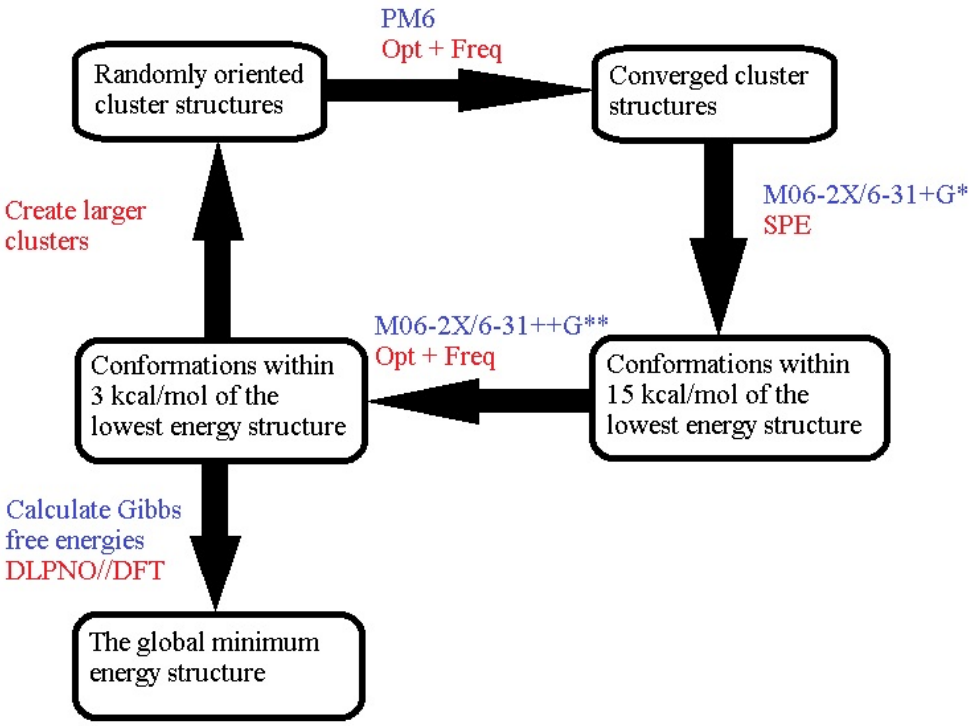

Figure 7: A schematic picture of cluster sampling technique used in **Papers III–VI** and **VIII**.

Multiple Conformers

The validity of the assumption, that global minimum energy structure can be used to approximate the properties of a multitude of clusters, has been tested, for example, by Temelso *et al.* and Partanen *et al.*[182,183] Temelso and co-workers have studied the sulfuric acid hydrates containing up to six water molecules. Because hydrogen bonds are highly flexible, the energy barrier for transforming one conformer to another, for instance through the rotation of a bond, can be assumed to be low. Therefore, one can expect that an ensemble of different conformers would be present for each cluster, referred to as global anharmonicity. The contribution of different conformers has been taken into account by using the Boltzmann averaged reaction enthalpy and Gibbs free energy for each sulfuric acid hydrate over the low-energy conformers. By accounting for multiple conformers the ΔG value is slightly larger than in the case of only

global minimum energy structure, meaning that Boltzmann averaging makes the cluster less stable.

Partanen *et al.* has showed that the calculation of thermal average is over all the relevant conformers leads to erroneous results even at the qualitative level, as the incorporation of higher energy conformers leads to an increase in the Gibbs free energy value. Actually, the presence of multiple conformers increases the number of energy levels and microstates, which means that ΔG must decrease. Partanen *et al.* has accounted for multiple conformers using statistical mechanics and compared the results with Boltzmann averaged energies. They found that the average difference in Gibbs free energies is 0.78 kcal/mol for the sulfuric acid hydrates reported by Temelso *et al.* This implies that when a large number of low-lying conformers is present, global anharmonicity has a larger impact on the thermochemical properties than local anharmonicity (vibrational anharmonicity see Section 3.4). Due to the large Gibbs free energy differences between different conformers, however, it is of utmost importance to identify the correct global minimum energy structure. In practice, this means that one conformer can be used in clustering studies, but a comprehensive conformational search must be done in order to locate the global minimum energy conformer.[183]

3 Quantum Chemical Methods

Starting from Dirac's equations, which offer the most complete description of N-electron systems,[184] a large number of approximations have to be introduced until one arrives at usable practical quantum chemical methods. It is often applicable to neglect or treat relativistic effects only approximately, and focus on the electronic ground state — allowing the use of an exact non-relativistic time-independent Schrödinger equation:[185]

$$\hat{H}\Psi(R,r) = \left(\hat{T}_{\mathrm{e}} + \hat{T}_{\mathrm{n}} + \hat{V}_{\mathrm{ne}} + \hat{V}_{\mathrm{ee}} + \hat{V}_{\mathrm{nn}}\right)\Psi(R,r) = E\Psi(R,r), \tag{8}$$

where E is the ground state energy of the system and $\Psi(R,r)$ is the corresponding wave function depending on the position of all nuclei (R) and electrons (r). The Hamiltonian $\hat{H}$ describes the potential and kinetic energy of the system, *i.e.*, the kinetic energy of the electrons $\hat{T}_{\mathrm{e}}$ and nuclei $\hat{T}_{\mathrm{n}}$ as well as the electrostatic interactions between nuclei and electrons $\hat{V}_{\mathrm{ne}}$, electrons themselves $\hat{V}_{\mathrm{ee}}$, and nuclei themselves $\hat{V}_{\mathrm{nn}}$.[186] In many cases it is recommended to apply the Born–Oppenheimer (BO) approximation where the motion of the nuclei is assumed to be negligible compared to the motion of the electrons.[187] This is justified because light

electrons move significantly faster than heavier nuclei and almost instantaneously respond to any changes in the relative position of the nuclei. Consequently this permits the separation of the Hamiltonian into its electronic and nuclear parts, where the electronic part can be solved separately while keeping the nuclear positions fixed. The mathematical description is known as the stationary N-electron Schrödinger equation

$$\hat{H}_{\mathrm{BO}}\Psi = \sum_{i=1}^{N}\left(-\frac{1}{2}\nabla_i^2 - \sum_{j=1}^{M}\frac{Z_j}{|\vec{r_i}-\vec{R_j}|} + \frac{1}{2}\sum_{i\neq j}^{N}\frac{1}{|\vec{r_i}-\vec{r_j}|}\right)\Psi, \tag{9}$$

where the BO Hamiltonian operator $\hat{H}_{\mathrm{BO}}$ includes all the physics necessary to describe the N-electron system. The N-electron wave function Ψ depends on N discrete spin variables and $3N$ spatial variables and it completely describes any non-relativistic N-electron system. Solving this equation yields an eigenvalue spectrum whose lowest eigenvalue corresponds to the ground state energy of the given system.[186] Each nuclear configuration yields a different set of energy levels, in fact, the eigenvalues change continuously with the nuclear configuration. However, there is no analytical solution for many-electron systems since the two-electron operator cannot be factorized, and thus more approximations are needed.

The general structure of the N-electron wave function is not fully known. According to the Pauli principle, since electrons are fermions they cannot occupy the same quantum state within a quantum system simultaneously and the wave function has to be antisymmetric with respect to interchange of electrons.[188] Generally speaking, same spin electrons avoid each other. This kind of electron correlation movement is called Fermi correlation. Typically in electron correlation methods, the total wave function $\Phi_{(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)}$ is constructed from a finite set of one-electron wave functions ϕ_N, in such a way that the Slater determinant fulfils the requirement of the Pauli principle:

$$\Phi_{(\mathbf{r}_1,\mathbf{r}_2,\ldots,\mathbf{r}_N)} = \frac{1}{\sqrt{N!}}\begin{vmatrix} \phi_1(\mathbf{r}_1) & \phi_2(\mathbf{r}_1) & \ldots & \phi_N(\mathbf{r}_1) \\ \phi_1(\mathbf{r}_2) & \phi_2(\mathbf{r}_2) & \ldots & \phi_N(\mathbf{r}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\mathbf{r}_N) & \phi_2(\mathbf{r}_N) & \ldots & \phi_N(\mathbf{r}_N) \end{vmatrix}. \tag{10}$$

It should be noted that the introduced approximations do not always work. The Born–Oppenheimer approximation will break down is some cases, for instance, when the energy gap between the ground state and the excited state is small, such as in metal surfaces, or there is coupling of more than one BO potential energy surfaces (PES), such as in photochemical reactions.[189–191] The time-independent Schrödinger equation leads to stationary states, which

do not change with time.[192] However, in the study of physical phenomena for which the potential energy of the system explicitly depends on time, treating time dependence is crucial.[193,194] For example, in spectroscopy when the system interacts with electromagnetic radiation, the excitations depends on time and thus the energy becomes time-dependent. For heavy elements the non-relativistic approximation breaks down.[195] This is because of a large positive nuclear charge makes core electrons move fast, close to the speed of light, and the relativistic mass of $1s$ electrons increase. The Bohr radius (which is the most probable distance between the nuclear and the electron) is inversely proportional to the electron mass, and thus the orbital contracts, which also affects the other orbitals.[196] Generally, the s and p orbitals contract and d and f orbitals expand. For instance, the colours of silver and gold can be traced back to the energy difference between the $(n-1)d$ and ns orbitals in the atom.[197] This transition is in the ultraviolet region for silver, giving the metallic luster. For gold the transition is in the visible region, however, only when relativistic effects are taken into account. The non-relativistic Schrödinger equation predicts gold to be of the same colour as the silver.[198] Understanding the limitations of the applied approximations is extremely important in quantum chemistry, especially because more approximations need to be introduced.

3.1 Basis Sets

In all quantum chemical methods described in this thesis, the electrons are placed in molecular orbitals (MOs) in terms of suitable basis functions.[199] A basis set is defined as a set of functions used to create the MOs, which are expanded as a linear combination of atomic orbitals (LCAO) with coefficients c_r to be determined

$$\psi = \sum_r c_r \chi_r. \tag{11}$$

The atomic orbitals (AOs) χ_r used in the expansion (11) constitute the basis set for the calculation. Usually the functions used to create the MOs are AOs centred on atoms, but they can also be centred on bonds or on lone electron pairs.[199] Typically calculations are performed with a finite set of basis functions, which can be Slater or Gaussian types. Slater type orbitals (STOs) are similar to the eigenfunctions of the hydrogen-like atom and they have a direct physical interpretation, however, they are computationally challenging since most of the required integrals must be calculated numerically.[200] Gaussian type orbitals (GTOs), in contrast, are computationally more convenient, but a single GTO gives a wrong physical

description for a hydrogen-like atom since it decays too fast from the nucleus and does not exhibit a cusp at the nucleus.[201] Hence, a common solution is to approximate STOs as a linear combination of Gaussian functions, which yields a more accurate description of the electron density near the nucleus than a single GTO, but in a computationally easier form than STO.[202] The form of Gaussian functions is

$$\chi^{\text{GTO}} = N(x-X)^k(y-Y)^l(z-Z)^m e^{-\zeta(r-R)^2}, \tag{12}$$

where N is the normalization constant and ζ is the exponent providing the radial extent of the function. The center of such a primitive function is $R(X, Y, Z)$, which is typically a nuclear position. The sum of k, l, and m defines the angular momentum, *i.e.*, the orbital subshell s, p, d, etc.[188]

Today, there are hundreds of basis sets composed of GTOs. The smallest possible basis set is called a minimal basis set, in which each atomic orbital in a Hartree–Fock calculation (including unoccupied orbitals) is described with a single basis function (see Section 3.2.1 for Hartree–Fock method).[186] For instance, each atom in the first row of the periodic table has a minimal basis set of two s-type functions and three p-type functions. The minimal basis sets are not flexible enough for accurate representation of the orbitals. The addition of multiple basis functions to describe an AO increases the computational accuracy, however, the cost of computation as well. To optimize the computational effort, *i.e.*, getting the maximal accuracy with a minimal computational cost, the core and valence electrons should be treated separately.[203] Core AOs are relatively independent of the chemical environment, meaning that a single basis function is often enough for their description. Valence AOs instead participate in the chemical bonding, thus requiring a more flexible description, namely multiple basis functions corresponding to each valence atomic orbital. These basis sets are called valence double-ζ, triple-ζ, quadruple-ζ, etc., indicating how many basis functions are used to treat valence AOs.[170]

In molecular environments, orbitals become distorted from their atomic shapes, meaning that they are polarized. To describe the effects of polarization functions having one additional node, *i.e.*, an angular moment one greater than the valence space, should be added.[204] Polarization functions increase the mathematical flexibility, which in turn allows molecular orbitals to adapt to their environment. Polarization functions are important for reproducing chemical bonding and they are required for quantitative calculations. When a system contains electrons which are not localized close to nuclei, diffuse functions are needed.[205] These functions have small exponents and they decay very slowly with the distance from the nucleus, meaning

that the probability to find an electron far away from the nucleus rises. Diffuse functions are necessary for systems which contain loosely-bound electrons or when long-range interactions are important, for example, anions, hydrogen bonds, dipole moments, and polarizabilities. The addition of diffuse functions is computationally very expensive.[170] It has been demonstrated that partial augmentation often yield as good results as the fully augmented basis sets, at significantly lower computational cost (see **Paper II**).[206]

Basis Set Superposition Error

When calculating weak molecular interactions such as non-covalently bound dimer interaction energies, basis functions from one molecule can help compensate for the basis set incompleteness on the other molecule in the dimer.[207] Hence, the dimer will be artificially too stable, and the difference is known as the basis set superposition error (BSSE).[208] In the limit of a complete basis set, the BSSE will be zero, but in practice this requires very large basis sets. An approximate way of assessing BSSE is the counterpoise (CP) correction. The Boys and Bernardi formula for the CP corrected binding energy of AB dimer formed from fragments A and B is:[209]

$$\Delta E^{\mathrm{CP}}_{\mathrm{binding}} = \Delta E_{\mathrm{binding}} + E^{\mathrm{A}}_{\mathrm{AB}}(\mathrm{A}) + E^{\mathrm{B}}_{\mathrm{AB}}(\mathrm{B}) - E^{\mathrm{AB}}_{\mathrm{AB}}(\mathrm{A}) - E^{\mathrm{AB}}_{\mathrm{AB}}(\mathrm{B}), \tag{13}$$

where $\Delta E_{\mathrm{binding}}$ is the uncorrected binding energy of AB dimer, $E^{\mathrm{A}}_{\mathrm{AB}}(\mathrm{A})$ and $E^{\mathrm{B}}_{\mathrm{AB}}(\mathrm{B})$ are energies for monomers A and B calculated with the geometry they have in the dimer, and $E^{\mathrm{AB}}_{\mathrm{AB}}(\mathrm{A})$ and $E^{\mathrm{AB}}_{\mathrm{AB}}(\mathrm{B})$ are energies for monomers A and B calculated at the dimer geometry and with the full dimer basis set.

Figure 8 presents an example of BSSE in the case of a sulfuric acid–dimethylamine complex. Binding energies are calculated using DLPNO–CCSD(T) method with aug-cc-pVxZ basis sets,[210–212] where the cardinal number x=2–5, with and without CP correction (see Section 3.2.8 for DLPNO–CCSD(T)). The BSSE is the difference between uncorrected and CP corrected values. As can be seen from Figure 8, BSSE decreases with higher cardinal number, but even using aug-cc-pV5Z the values do not coincide (see **Paper VIII** for more discussion).

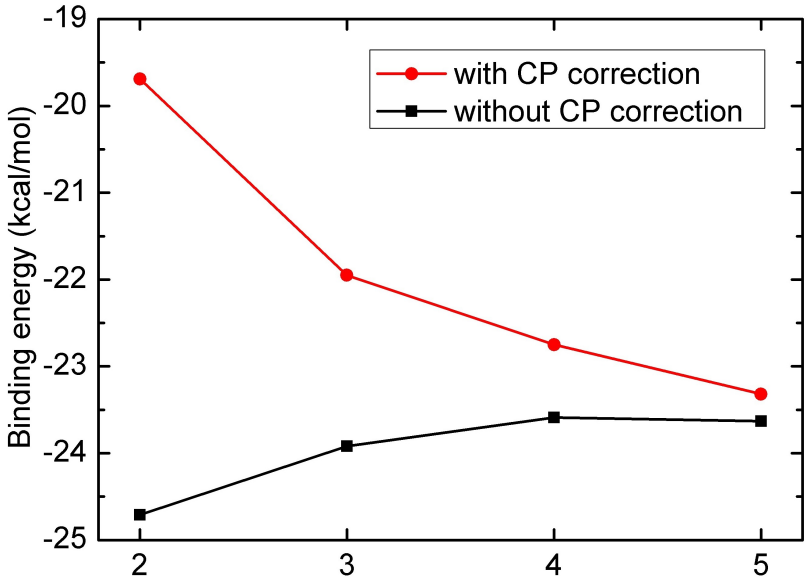

Figure 8: The DLPNO–CCSD(T)/aug-cc-pVxZ binding energies (kcal/mol) for sulfuric acid–dimethylamine complex with and without counterpoise (CP) correction as a function of basis set cardinal number x=2–5.

3.2 Wave Function Theory

In wave function based methods, both the accuracy and the computational cost are highly dependent on the level of theory. Most wave function methods converge towards the exact energy and all properties simultaneously. The accuracy is dependent on the treatment of electron correlation, the basis set size, and the Hamiltonian used. The computational cost is often referred to as scaling with respect to system size. The system size N means simply the number of basis functions, which is dependent on the number of atoms and the flexibility of the solution, *i.e.*, the basis set used. The time taken by algorithm T can be written using the scaling law as

$$T = aN^b, \tag{14}$$

where a is the prefactor and b is the scaling exponent. The computational bottleneck in wave function based methods is the high scaling exponent, which restricts the most accurate methods to the smallest systems. The holy grail in developing new and efficient algorithms is a linear scaling together with a small prefactor.

3.2.1 Hartree–Fock

The starting point for most approximate quantum chemical wave function theory is the Hartree–Fock (HF) method, where each electron interacts only with the average electrostatic field of all other electrons.[213] The orbitals are optimized according to the variational principle and the formed Slater determinant would be, in the limit of an infinite basis set, an exact solution of the Hartree–Fock Hamiltonian. Since the wave function is built by spin orbitals depending on one electron, only the same-spin electrons can interact with each other through the Slater determinant, which means that the electron–electron repulsion term is not properly treated. The Hartree–Fock wave function only accounts for the Fermi correlation (repulsion between same spin electrons) but not the Coulomb correlation (Coulomb repulsion between negative charges) because the instant electron–electron interaction is missing.[214,215] Therefore, HF method convergences reasonably fast to the complete basis set limit. Typically, this scheme can recover more than 99% of the total electronic energy, however, being still far away from chemical accuracy (defined as within 1 kcal/mol of the exact solution).[216] The missing 1% in the HF method has a crucial impact on chemistry, relating to the ability of an electron to respond to the other electrons motion. Classically speaking, each electron avoids other electrons as much as possible, and with a certain probability move into the virtual orbitals if they come close to each other. This type of electron correlation is neglected in the Hartree–Fock method. Therefore, the electrons are located too closely to each other meaning that bond lengths are too short, which leads to overestimation of vibrational frequencies and activation energies.

In addition, dispersion interaction is completely missing in the HF approach. Dispersion interaction is a long-range electron correlation effect, which is often very important in the case of describing weakly-bound molecular clusters. Non-bonded fragments can interact, depending on the distance, repulsively or attractively. In the HF method, the repulsion is taken into consideration in the self-consistent field (SCF) stage. However, when describing dispersion forces, electron correlation is necessary to take into account. Each molecule has always a small, variable dipole moment due to the rapid movement of electrons. This non-permanent dipole moment creates attraction by polarizing surrounding molecules, and the HF method does not account for this type of dispersion forces.[217] The dispersion forces also affect the stability of molecules, for example, the stability of branched alkanes over linear ones is due to electron correlation. Because of this, the Hartree–Fock method is not recommended when the electron correlation between reactants and products changes, since it does not treat the reactants and products equivalently. To reach chemical accuracy — or

in some cases even a proper physical description — the treatment of electron correlation is mandatory. Therefore, numerous approaches are developed upon the HF solution to cover at least part of the correlation energy.[170]

3.2.2 Electron Correlation

A strong feature of Hartree–Fock method is that it is variational, meaning that the HF energy always corresponds to an upper limit to the ground state energy and the result can be improved systematically by including more electron correlation. The correlation energy can be divided into three different terms: 1) Fermi correlation, which arises from the Pauli antisymmetry of the wave function and, as mentioned earlier, is taken into account already in the Hartree–Fock method, 2) dynamic correlation, which is associated with the instantaneous correlation among the electrons arising from their mutual Coulomb repulsion, and 3) static correlation, which arises from near-degeneracy of electronic configurations.[186] Therefore, the missing correlation energy E_{corr} consists of parts 2) and 3) and it is defined as the difference between the exact energy E_{exact} and the single-determinant HF energy E_{HF} in the same basis as

$$E_{\text{corr}} = E_{\text{exact}} - E_{\text{HF}}. \tag{15}$$

All approximations in the solution of the Schrödinger equation should be unambiguous and precisely defined as well as improvable in a systematic fashion. There exists several different approaches to construct an approximative electronic wave function; starting from the simplest Hartree–Fock model, where the wave function is represented by a single Slater determinant, up to the most complex full configuration interaction (FCI) model, where the wave function is represented as a variationally determined superposition of all determinants in the N-electron Fock space.[218] Between these levels, there are a large amount of different approaches, which employ additional contributions of electron correlation with variable accuracy and computational cost. It should be noted that none of the methods is suitable to all systems — therefore, one is likely to spend a lot of quality time to find a theoretical level to apply for a given problem.

The frozen core approximation (FCA) is usually used as a default in electronic correlation calculations.[219] It is based on the fact that chemical reactions happen mainly via valence electrons, and that the core electrons are inactive from a chemical point of view. Therefore, the lowest-lying molecular orbitals are constrained to remain doubly-occupied in all configurations, which reduces the number of configurations. A justification for the FCA is that the core

electrons of an atom are less sensitive to their environment than the valence electrons. Thus the error introduced by freezing the core orbitals is nearly constant for molecules containing the same types of atoms. Because of this error cancellation, the core electrons can be left inactive in a correlation treatment when handling relative energies. However, core electrons do affect the absolute energies, and thus FCA is not a good approximation for total electronic energies. When one need to include the core electrons into correlation treatment, specialized core-correlation designed basis sets are needed.[186]

Both dynamic and static correlation effects can be taken into account by mixing in Slater determinants for more electron configurations Φ_i to the HF one Φ_0

$$\Psi = c_0\Phi_0 + \sum_i c_i\Phi_i, \tag{16}$$

where c_0 is the coefficient for HF solution and c_i is the coefficient for excited electron configuration determinant. If c_0 is assumed to be close to 1 and a large number of excited determinants Φ_i are added, each of which is assumed to give only a small contribution, then the method primarily treats dynamic correlation. On the other hand, if it assumed that there are just a few excited determinants with coefficients close to the reference one, then the method primarily treats static correlation.

Static correlation is a long-range correlation effect, which deals with only few, but crucial determinants. It needs to be taken into account in situations where multiple determinants are required to cover the electronic structure of a state, *i.e.*, when a single Slater determinant yields a qualitatively wrong description of the electron configuration.[170] Generally speaking, static correlation is not needed for closed-shell systems at their equilibrium geometries, but it might be necessary when stretching a bond or bending an angle. For example, when twisting the double bond of ethylene, at a ninety degrees twist angle the systems behaves like a biradical, and hence the use of a multireference method is needed.[220] A common way to take static correlation into account is the complete active space self-consistent field (CASSCF) method.[221] It is a special form of a multiconfigurational SCF method and can be thought of as an extension of the Hartree–Fock method. In the CASSCF method a FCI is performed for an active space, which is constructed from n electrons in m orbitals. CASSCF is a powerful method to study static correlation effects, however, calculations are fairly complex and ultimately require a lot of insight from the user in order to be successful. It should be noted that CASSCF calculations are not designed to provide accurate absolute energies. The purpose of a CASSCF computation is to provide a qualitatively correct wave function for a good starting point of dynamic electron correlation calculation. For example,

the Complete Active Space Second Order Perturbation Theory (CASPT2) can be used to add dynamic correlation by perturbation theory on top of the CASSCF solution.[222] The number of configurations increases factorially with the size of active space, and therefore the limit of feasibility is roughly around 14 active orbitals or about one million configuration state functions (CSFs) in the active space. For large active spaces, a new promising approach is the Density Matrix Renormalization Group (DMRG) which yields results close to the CASSCF method.[223,224] Since multireference methods have not been used in any of the papers included in this thesis, they will not be treated further here.

Dynamical correlation has a significant role for systems studied in this thesis. It is a short-range correlation effect, which refers to capturing the effect of the instantaneous electron repulsion by allowing the electrons to respond rapidly to the movement of the other electrons. It is based on the assumption that the single-determinant description is qualitatively correct, and thus the correction can be done by throwing in lots and lots of determinants with very small weight each. There are three main methods to include dynamic correlation within wave function theory: configuration interaction (CI),[225] Møller–Plesset (MP) perturbation theory,[226] and coupled cluster (CC) theory.[227]

3.2.3 Configuration Interaction

Configuration interaction is a variational method in which the trial wave function is given as a linear combination of determinants with the expansion coefficients determined by requiring the energy to be a minimum

$$\Psi_{\mathrm{CI}} = a_0\Phi_0 + \sum_S a_S\Phi_S + \sum_D a_D\Phi_D + \cdots = \sum_{i=0} a_i\Phi_i. \tag{17}$$

The excited electron configurations are obtained by exciting electrons from occupied orbitals to virtual orbitals so that the total spin of the system does not change. The molecular orbitals (MO) used for building the excited Slater determinants are taken from HF calculations and held fixed. The created wave function is then optimized with respect to the total energy and the new energy is lower than the HF energy. If all possible excitations are taken into account (FCI), the obtained energy would be exact in the given basis set.[228] However, the FCI method is not feasible for all but the very smallest systems because of the number of excited Slater determinants increases factorially with the system size (N^N), and therefore, the amount of excitations is typically truncated. The nomenclature for excitations is dependent on how many

electrons are excited in all possible combinations, such as single is S, double is D, triple is T, and so on.[225]

According to the Brillouin's theorem, the CIS method does not improve the HF result as all matrix elements between the HF wave function and singly excited determinants are zero.[229] However, although the ground state is equal, single excitations may be used as approximations to excited states and they allow the CI wave function to relax the MOs. Thus the lowest model which yields improvement over HF results is CID. There is only a marginal computational increase in effort of CISD over CID since the number of singly excited determinants is relatively small. Single excitations enter the wave function indirectly as they have non-zero matrix elements with the doubly excited determinants. Moreover, single excitations do affect the electronic charge distribution and therefore properties such as the dipole moment and polarizability. The CISD method scales N^6 and it is only CI method applicable for a wide variety of systems. The inclusion of the triply excited determinants, CISDT, and the quadruply excited determinants, CISDTQ, increases the scaling to eight and tenth power, respectively. While the FCI method is size consistent, all truncated CI methods suffer from a huge drawback by not being size consistent, which results in a non-physical scaling of the error with the system size.[170] Although many attempts have been made to partially repair this problem, for instance the Davidson correction, the CI methods are not of great use for ground states anymore.

3.2.4 Møller–Plesset Perturbation Theory

The idea behind perturbation theory is that the Hartree–Fock solution only differs slightly from the desired exact solution, and thus the solution may be improved by adding electron correlation as a small perturbation. For instance, when the HF wave function contains considerable multi-reference character, single determinant MP theory will display weak convergence. Therefore, for systems containing nearly degenerate orbitals, the reference wave function might be qualitatively wrong, and then MP methods are not recommended.[217]

After the Schrödinger equation for the reference Hamiltonian operator is solved, the Schrödinger equation for the perturbed system can be written as

$$\hat{H}\Psi_n = \left(\hat{H}_0 + \lambda\hat{H}'\right)\Psi_n = E_n\Psi_n, \tag{18}$$

where the parameter λ takes values from zero to one, representing the extremes of the unperturbed and fully perturbed systems, respectively. $\hat{H}_0$ is the unperturbed Hamiltonian

and $\lambda\hat{H}'$ is the perturbation that describes electron correlation. The first order correction is the average of the perturbation operator over the zeroth order wave function and is already included in the HF energy. Therefore, the lowest method which yields improvement over HF results is MP2. The second order correction to the energy is

$$E^{(2)} = \sum_{m \neq n} \frac{|\langle \Psi_m^{(0)} | \hat{H}' | \Psi_n^{(0)} \rangle|^2}{E_n^{(0)} - E_m^{(0)}}. \tag{19}$$

Here is used the Dirac bracket notation, which is explained for instance in Ref. 188.

MP2 represents the simplest single reference correlation method and it is often a significant improvement over Hartree–Fock, recovering typically 80–90% of the correlation energy.[186] It scales as N^5, being computationally the most economic way to include electron correlation. A drawback of the MP method is the non-variational nature, which means that higher order corrections do not necessary yield better results, and the energy may be lower than the exact energy.[230] It turns out that the MP2 method typically overshoots the correlation effect, such as overestimating π–π interactions and overbinding molecules.[231] However, it describes hydrogen bonds well. Another advantage is that MP perturbation theory (unlike other perturbation methods) is size consistent.

3.2.5 Coupled Cluster

One of the most mathematically elegant techniques for estimating the electron correlation energy is coupled cluster (CC) theory.[232] It has become a cornerstone of modern *ab initio* computational chemistry. The coupled cluster method represents a well-defined and systematic approach to include dynamical correlation effects. In CC theory, an exponential cluster operator is used to account for electron correlation as

$$\Psi_{\text{CC}} = e^{\hat{T}} \Phi_0. \tag{20}$$

The non-Hermitian exponent operator is expressed as the sum of all excitation operators $\hat{T}_n$

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots + \hat{T}_n, \tag{21}$$

where $\hat{T}_1$ contains single excitations, $\hat{T}_2$ double excitations, and so on. The exponential term in Equation (20) can be written as

$$e^{\hat{T}} = 1 + \hat{T} + \frac{\hat{T}^2}{2!} + \frac{\hat{T}^3}{3!} + \ldots = \sum_{k=0}^{\infty} \frac{\hat{T}^k}{k!}. \tag{22}$$

The coupled cluster method provides the exact solution to the time-independent Schrödinger equation when all excitations are included, but in practice, the series expansion is truncated to include only specific electron excitation operators.[217] The level of truncation is indicated in a similar way by letters as in the CI method. The simplest method which improves HF result is CCD, which contains only double excitations. In the CCD method, the Taylor expansion of the exponential function gives

$$\Psi_{\text{CCD}} = \left(1 + \hat{T}_2 + \frac{\hat{T}_2^{\,2}}{2!} + \frac{\hat{T}_2^{\,3}}{3!} + \dots\right)\Phi_0. \tag{23}$$

The first two terms in the parenthesis in Equation (23), $1 + \hat{T}_2$, are equal to CID method, but the remaining terms involve products of excitation operators. This means that the square of $\hat{T}_2$ generates quadruple excitations, the cube of $\hat{T}_2$ generates hextuple excitations, and so on. Even though the excitations are truncated to a finite order, higher excitations are still included, and this corrects the failure of CI method, *i.e.*, CC methods are size consistent at any given order. Moreover, due to the exponential parametrization of the wave function, a certain amount of orbital relaxation is introduced, which greatly enhances the stability in multi-reference cases. Unfortunately, the truncated CC theory is not variational.[199]

In practice, the increase in accuracy of including single excitations in addition to doubles is worth the cost, and this leads the CCSD method with N^6 scaling.[233,234] Although considering only singles and doubles, CCSD still gives the exact results for two-electron system within a given basis. The CCSD method typically recovers 90–95% of the correlation energy. By itself CCSD is not a highly accurate method, for instance, it underestimates π–π interactions remarkably. However, it represents a feasible and robust electronic structure approach that is an excellent starting point for the calculation of the remaining correlation energy and molecular properties, at least for single-reference systems.

The CC2 method is an approximate second-order coupled cluster model which was constructed with emphasis on the calculation of molecular properties rather than on total energies.[235] It should be noted that particularly important for molecular properties are the singles which give an approximate orbital relaxation. Consequently CC2 approximates the CCSD doubles equations to the form occurring in first order but with the singles retained to provide an approximate description of orbital relaxation. For ground states, CC2 allows for a relaxation of connected single excitations, which are neglected in MP2, but this does not lead to a systematically improved accuracy.[236] Because the doubles are correct to the first order, the CC2 energy is correct to the second order and is expected to be of similar quality as the MP2

energy.[236,237] The CC2 method scales N^5 like MP2, but in contrast to MP2, excitation energies and transition moments can be obtained in CC2.[235] CC2 has been develop for calculations on excited states and was never intended for ground state calculations; however, Kurtén *et al.* and Ortega *et al.* have utilized the method for the ground states of reactions and clustering in the atmosphere (see Section 3.5.1 and **Paper I** for discussion).[238,239] Inclusion of connected triple excitations to the CCSD model defines CCSDT which scales N^8, and hence it is not common in practical use. The CC3 method is based on approximating the CCSDT model and it scales N^7.[240] The hierarchy for coupled cluster models can be expressed as CC2, CCSD, CC3, CCSDT, where the computational effort increase by a factor of N in each level. Also energies and properties increase in accuracy at each step with a convergence towards the FCI solution that is unique to the coupled cluster theory.[232]

To reach chemical accuracy, the most successful method is CCSD(T),[241] in which single and double excitations are included with a full treatment and an estimate to the connected triples contribution is calculated non-iteratively using perturbation theory.[233] In general, the inclusion of perturbative triplets slightly overestimates the triples correction, and does so by an amount about equal to the ignored quadruples. Therefore, CCSD(T) usually yields relative energies very close to the FCI limit, and indeed, it has come to be the effective gold standard for single-reference calculations.[242] Due to the very high computational expense and the poor scaling, N^7, it is unsuitable for routine applications and restricted only to small systems (see **Paper I**). However, it is widely used for benchmarking more approximate methods as well as studying ground state properties of small molecules. If higher quality results than obtained from CCSD(T) are required, also other error sources, such as non-relativistic, frozen core, and Born–Oppenheimer approximations, must usually be accounted for.[199] But as one might guess, the most accurate quantum chemical methods are limited to the very smallest systems containing a maximum of a few atoms or a few tens of electrons. Due to the computational restrictions, it is not surprising that a huge amount of effort is used to develop approximative methods which obtain a major increase in speed without a significant loss of accuracy.

3.2.6 Resolution of the Identity

When one desires to apply calculations to large molecules, the full integral transformation from the atomic orbital (AO) to the molecular orbital (MO) basis becomes extremely memory intensive and the most time-consuming part of the calculation.[243] The resolution of the identity (RI) approximation (known also as density fitting) allows performing the AO to MO transformation and some of the contractions with lower-index integrals instead of four-index integrals and removes the need to save any four-index integrals on disk.[244] It effectively reduces the computational requirements of the transformation step without a notable loss of accuracy. Resolution of the identity equation connects every complete basis set to the identity operator on that space. First the product of two basis functions $|\alpha\beta\rangle$ is expanded in an auxiliary basis set P as

$$|\alpha\beta\rangle = \sum_{P}^{M_{\text{aux}}} c_P P \tag{24}$$

and then four-index integrals are replaced by products of lower-index integrals as

$$\langle\alpha\beta|\gamma\delta\rangle = \sum_{PQ}^{M_{\text{aux}}} c_P c_Q \langle\alpha\beta|P\rangle\langle P|Q\rangle^{-1}\langle Q|\gamma\delta\rangle, \tag{25}$$

where $\langle P|Q\rangle$ are two-center and $\langle\alpha\beta|P\rangle$ are three-center integrals. The expansion coefficients c_P are determined by minimizing the fitting error, for example, using a Coulomb metric. Here is used the second quantization, which is explained for example in Ref. 170.

The number of three-center integrals is much smaller than four-center integrals, which reduces the formal scaling, however, actual timing shows that the total computational cost is reduced by approximately an order of magnitude.[245] This means, in fact, that the RI approximation reduces only the pre-factor. In addition, the calculation time per three-center integral is up to a factor of ten smaller than four-center integral. Some of the necessary integrals are fitted in an auxiliary basis set, which means from a user's point of view that an auxiliary basis set also must be chosen. The RI-fitted Coulomb energy always undershoots the exact Coulomb energy, however, in relative energies the error cancels out with reasonable auxiliary basis sets.[246] Hence, the RI approximation reduces the integral calculation cost remarkably by only introducing minor errors to the calculation and the use of it is strongly recommended.[247]

3.2.7 Explicit Correlation

One common problem in electron correlation methods is a frustratingly slow convergence towards the exact energy with respect to the basis set size.[248] This originates from the usage of antisymmetrized products of Slater determinants to construct two-electron (and higher-rank) basis sets.[249] Slater determinants fail to model the exact wave functions at short interelectronic distances. The difficulties of describing short-range dynamical correlation are related to the singularities in the Hamiltonian which give rise to the Coulomb cusp in the wave function.[250] The Coulomb cusp is caused by the instant electron–electron interaction and it cannot be described properly with smooth functions. The theory of explicit correlation can be used as an alternative to basis set extrapolation. In basis set extrapolation techniques, one performs calculations with lower cardinal number basis sets and extrapolates the results to correspond complete basis set limit (CBS).[251] Basis set extrapolation requires an arbitrary fitting, whereas explicitly correlated methods solve the convergence problem by using a wave function that depends explicitly on terms of the interelectronic distances. Therefore, explicit correlation technique is theoretically better justified than basis set extrapolation.[252]

However, for many-electron systems an explicit introduction of the interelectronic distance coordinate directly into the wave function is not an easy computational problem. Adding terms to the wave function approach that contains the interelectronic coordinates explicitly speeds up the basis set convergence.[253] Thus explicitly correlated methods yield near basis set limit results for correlation energies in conjunction with significantly smaller orbital basis sets. The terms added can be linear in the interelectronic distances, which is denoted R12 methods.[254] At long range, a linear correlation factor yields an unphysical behaviour and hence different correlation factors have been tested. Currently, the most popular choice is the F12 method, in which an exponential function of the interelectronic distances r_{12} is used. The correlation factor in the F12 method is taken to be a simple Slater function

$$F(r_{12}) = \exp(-\beta r_{12}), \tag{26}$$

where β is the length-scale parameter accounting for the strength of the interelectronic interaction. In practice, the Slater function is approximated by a linear combination of Gaussians. For example, the CCSD-F12 wave function[255] can be expressed as

$$\Psi_{\text{CCSD-F12}} = e^{\hat{T}_1+\hat{T}_2}\Psi_0, \tag{27}$$

where single excitations are

$$\hat{T}_1 = t_a^i \hat{E}_{ai} \tag{28}$$

and double excitations are

$$\hat{T}_2 = T_{ab}^{ij}\hat{E}_{ai}\hat{E}_{bj} + \tau_{\alpha\beta}^{ij}\hat{E}_{\alpha i}\hat{E}_{\beta j}. \tag{29}$$

Indices i refer to occupied orbitals, a to virtual orbitals, and α to a complete orbital basis set. t_a^i and T_{ab}^{ij} are the conventional singles and doubles amplitudes, respectively. In F12 theory, the correlation factor is used to approximate the additional amplitudes $\tau_{\alpha\beta}^{ij}$ as

$$\tau_{\alpha\beta}^{ij} = \langle\alpha\beta|\hat{Q}_{12}\hat{F}_{12}|kl\rangle T_{kl}^{ij}, \tag{30}$$

where $\hat{Q}_{12}$ is a projector which ensures strong orthogonality of the explicitly correlated terms to the HF reference function and to the conventional double excitations.

The disadvantage of the explicitly correlated approach is the emergence of numerous three- and four-electron integrals, which in practices restricts methods to small systems. Therefore, the many-electron integrals are calculated using the resolution of the identity approximation so that only two- and three-electron integrals remain. The faster convergence of explicitly correlated methods means that fewer high angular momentum functions are needed, and not surprisingly, the basis set requirements of such methods are different than regular electron correlation methods. In practise, this means that when using F12 methods one should specify a special orbital basis set (OBS) as well as a complementary auxiliary basis set (CABS) in order to control the error coming from the RI approximation.[256] The development of F12 basis sets is still in early stages. Currently, the cc-pVnZ-F12 basis sets (where n=2–5) are available for a restricted range of elements.[257,258] It should be noted that in contrast to what their names suggest, F12 basis sets are a fair bit larger than the corresponding cc-pVnZ basis sets. In the case of carbon, for example, the cc-pVDZ basis set contains six basis functions, whereas cc-pVDZ-F12 contains 12. Calculations performed at the CCSD(T)-F12/cc-pVDZ-F12 level of theory[259] yields results of CCSD(T)/cc-pVQZ quality.[260,261] This implies significantly faster basis set convergence with some additional computational cost. The extra computational effort necessary for F12 calculations is due the additional two-electron integrals. The computations can be strongly accelerated by means of the RI approximation utilized for the F12 part.[262] Generally, it is then recommended to use one cardinal number higher basis set for the RI approximation, for example, in conjunction with a cc-pVTZ-F12 basis set, one should apply a cc-pVQZ/C fitting basis.[263]

3.2.8 Local Electron Correlation Approaches

Introducing explicit correlation into the wave function greatly accelerates the basis set convergence, but unfortunately the scaling behaviour remains the same. Methods that aim at reduced scaling are based on the fact that the dynamic electron correlation is a short-range effect.[264] Notably, although correlation is a local effect it is not arbitrarily local. Pair correlation energies falls off as R^{-6} covering a region of space that extends over 2–3 chemical bonds (sometimes more in a conjugated π-system). The aim of local correlation methods is to reduce the unpleasant scaling with system size — ideally to (sub)linear — and to preserve the accuracy of wave function based correlation approaches.[265] This can be done only if the introduced error exploiting the locality is not spoiling the intrinsic accuracy of the method. In order to keep the desired accuracy, the method should cover at least 99.9% of the basis set correlation energy, *i.e.*, the correlation energy in a given basis.

Methods that exploit the locality of electron correlation can be divided loosely into two complementary classes: piecewise and direct local approaches. Piecewise methods use locality by dividing the molecule into subsystems and performing small calculations for them at the same time. These results are combined to estimate the total correlation energy. Piecewise local approaches are for instance the divide-and-conquer (DC) proposed by Yang and Lee[266,267] and developed by Li and Li,[268] divide-expand-consolidate (DEC) by Jørgensen *et al.*,[269,270] and clusters-in-molecules (CIM) by Li *et al.*[271,272] Advantage of these approaches is that they are highly efficient parallel and readily extended to properties.

Direct local correlation methods aim to achieve a truncation of the virtual space while performing the calculation on the entire system. Locality is used in the algorithm to avoid calculating terms close to zero or factors that are unity. Saebø and Pulay suggested spanning the virtual orbital space by projected atomic orbitals (PAOs).[264,273] Whereas conventional unoccupied orbitals span the entire molecule, PAOs are obtained by projecting the occupied orbital space out of the atomic orbitals and largely preserve the localized character of the atomic orbitals. Elaborating on the concept of PAOs, a variety of correlation methods have been developed and implemented. For example, Schütz and Werner demonstrated linear scaling local methods of LMP2,[274] LCCSD,[275] LCCSD(T),[276] and LCCSDT.[277] These local methods have revolutionized the applicability of reliable correlation methods to large molecular systems. In order to reduce the computational effort, Schütz and Werner have applied density fitting for local methods and developed DF-LMP2, DF-LCCSD, and DF-LCCSD(T).[278–280] A major benefit of the concept of correlation domains is that they allow

linear scaling with respect to CPU, disk and main memory. A major drawback, in turn, is that they need to be specified prior to a calculation, which results in limited error control. Also methods that implement orbital-specific virtuals (OSVs) are proposed.[281] OSVs are not pair-specific, but orbital-specific virtual orbitals, meaning that there is a correlating set of virtual orbitals specific for each occupied molecular orbital. Using OSV approaches it is possible to achieve linear scaling with respect to system size, and 92–97% of the triples correction can be covered.[282]

Already in the 1965, Edmiston and Krauss introduced pair natural orbitals (PNOs).[283,284] A few years later Meyer[285,286] and Ahlrichs *et al.*[287,287] continued with the concept of pair natural orbitals. However, only in recent years have PNOs been under active development.[288] The general idea of the PNO-based approach is to construct approximate natural orbitals that are specific for each electron pair. The PNO space for a given electron pair is local and located in the same region of space as the corresponding electron pair. This means that PNOs, in contrast to PAOs, are pair specific and lead to a very compact representation of the virtual space. The most severe bottlenecks of the PNO approach are the laborious integral transformations, which can be avoided by using the RI approximation. Local pair natural orbital (LPNO) methods, in which the internal space is spanned by localized internal orbitals, have been successfully applied for medium-size systems. The first LPNO method implemented was the coupled-electron pair approximation (CEPA).[289] Afterwards the CCSD method was developed within the LPNO framework.[290] The LPNO methods always make use of the RI approximation, which means from the user's point of view that an auxiliary basis set must be provided. The LPNO–CCSD method scales N^5 which restricts it for systems about 100 atoms, and therefore, it was subsequently redesigned to address its inherent scaling.[288]

DLPNO–CCSD(T)

The LPNO approach was further developed by combining the concepts of PNOs and PAOs, which led to the near linear scaling domain-based local pair-natural orbital CCSD (DLPNO–CCSD) method.[291] The addition of quasi-perturbative treatment of the triple excitations yielded the DLPNO–CCSD(T) method,[292] which allows calculations on molecules with hundreds of atoms or with nearly 9000 basis functions. The DLPNO–CCSD(T) method differs slightly from the corresponding canonical methods by employing the T_0 approximation for the evaluating of triples. In the T_0 approximation all off-diagonal Fock matrix elements are neglected, which yields a significant computational speed-up. Neese and co-workers demonstrated the very first CCSD(T) level calculations on an entire protein consisting of 644 atoms (Crambrin).[292] Figure 9 shows the scaling behaviour of conventional CCSD(T) and

[D]LPNO-approximated coupled cluster methods. All the [D]LPNO methods are implemented in the Orca quantum chemistry program.[293]

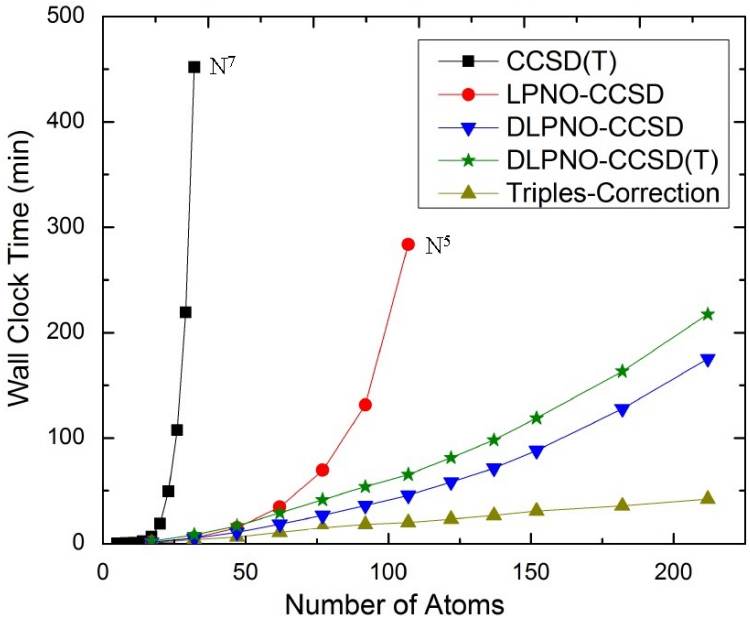

Figure 9: Scaling behaviour of the canonical CCSD(T), LPNO–CCSD, DLPNO–CCSD, and DLPNO–CCSD(T) methods for linear hydrocarbons. Figure adapted from reference 294.

It should be noted that various approximations are needed in order to reach near linear scaling. The theory of the DLPNO–CCSD(T) method has been described in detail by Neese and co-workers,[292] and here only the most important features are explained. The DLPNO method is essentially a black-box method, with three tunable parameters: T_{CutPairs} which controls a perturbative selection of significant pairs, *i.e.* which pairs are included in the CCSD iterations, T_{CutPNO} which is the PNO occupation number which largely controls the accuracy, *i.e.* the number of PNOs per electron pair, and T_{CutMKN} which defines the domain size for the local fit to the PNOs within the RI scheme. First the localization of the occupied orbitals is obtained from a single determinant reference wave function computation. The total electron correlation is given by the sum over electron pair correlation energies. In strong-pair approximation the strong and weak pairs are defined based on MP2 pair correlation estimates (T_{CutPairs}). The strong pairs enter the coupled cluster iterations, whereas for weak pairs only MP2 additive corrections to the total correlation energy are calculated (Figure 10).[288] The correlation virtual space of each strong pair consists of PNOs which are constructed from the MP2 pair densities. The MP2 correction is calculated when the occupation number of PNO is larger than the

threshold T_{CutPNO}. Then the required PNOs and integrals are expanded in terms of PAO domains, whose sizes are controlled by T_{CutMKN}. Default threshold values in Orca version 3 are $T_{\text{CutPairs}} = 10^{-4}$, $T_{\text{CutPNO}} = 3.33 \times 10^{-7}$, and $T_{\text{CutMKN}} = 10^{-3}$.

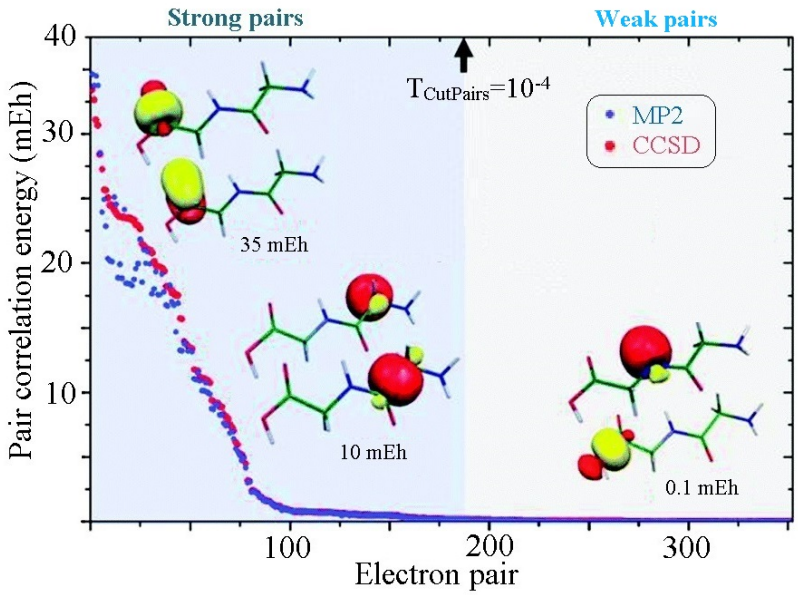

Figure 10: The electron pairs are divided into weak and strong pairs based on a local MP2 estimate of the pair correlation energy. The strong pairs are explicitly included in the coupled cluster procedure, whereas the weak pairs are added afterwards to the total correlation energy. Figure adapted from reference 288.

Benchmark studies have shown the DLPNO–CCSD(T) method to be a reliable and an affordable way to cover electron correlation for large molecules (see **Paper I**). Typically DLPNO–CCSD(T) covers 99.8–99.9% of the canonical coupled cluster correlation energy.[292] Since DLPNO does not require the user to adjust any parameters, it can be used in a black-box fashion. Gradients are not yet available for [DL]PNO methods, but they are under active development. For instance, Hättig and co-workers have presented a preliminary implementation of gradients for the PNO–MP2 method.[295] In addition, multireference (DLPNO–NEVPT2)[296] and explicitly correlated (DLPNO–MP2-F12)[297] methods are already implemented and further development is under process.

The increasing complexity of local correlation methods makes codes difficult to develop. A major challenge is the lack of robust software for handling sparse tensor operations. Neese and co-workers have introduced a sparse map infrastructure for dealing with sparse tensor

data structures and algorithms which appear in local correlation approaches.[298] Here all essential computational steps are achieved in a linear scaling fashion, for example, the integral transformation, PNO construction triples correction, and amplitude iterations. On the basis of the concept of sparse maps, a linear scaling DLPNO–CCSD(T) method is presented in 2016 and calculations with more than 20000 basis functions and 1000 atoms are reported.[299] The new DLPNO–CCSD(T) method is about seven times faster and uses four times less disk space than the old one. For instance, the single point energy calculation for the Crambrin protein using old DLPNO–CCSD(T) method with a def2-SVP basis set[300] takes 30 days of CPU time and 1.3 TB disk space, and the same calculation with linear scaling DLPNO takes only 5 days of CPU time and 0.4 TB disk space.[299] In addition, new DLPNO reduces the error in absolute correlation energies by approximately a factor of two compared to the old DLPNO method. In 2015, Liakos and co-workers defined three default thresholds, which are used to control the absolute desired accuracy.[301,302] Default thresholds are LoosePNO, NormalPNO, and TightPNO. Authors recommend to use LoosePNO for preliminary studies, NormalPNO for general thermochemistry and kinetics, and TightPNO for non-covalent interactions and conformational equilibria. NormalPNO and TightPNO provide relative energies within 1 kcal/mol of canonical CCSD(T) calculations.[301,302]

Figure 11 shows an example of sulfuric acid–ammonia complex binding energies calculated using old DLPNO with default criteria, new DLPNO with TightPNO criteria, and in comparison with canonical RI-CCSD(T)-F12 using different basis sets.[261] New DLPNO with TightPNO estimates binding energies to be on average 0.31 kcal/mol lower than the old DLPNO with default criteria, with a variation of 0.28 to 0.37 kcal/mol. All other DLPNO binding energies differ less than 0.7 kcal/mol compared to the highest level RI-CCSD(T)-F12/VQZ-F12, except when using a double-zeta basis set without diffuse functions. It can be noticed that DLPNO/aug-cc-pVTZ corresponds well to the canonical binding energies, and DLPNO/def2-QZVPP gives results close to DLPNO/aug-cc-pV5Z. Since DLPNO does not cover all of the correlation energy of canonical CCSD(T), it does not converge to the same value as canonical coupled cluster.[292] From a computational point of view this means that when choosing the basis set, user should decide whether to trust error cancellation to be systematic and use aug-cc-pVTZ or to get binding energies close to DLPNO basis set limit and utilize def2-QZVPP. Recent benchmark studies have shown that DLPNO/aug-cc-pVTZ does indeed consistently yield results in good agreement with canonical coupled cluster calculations.[90]

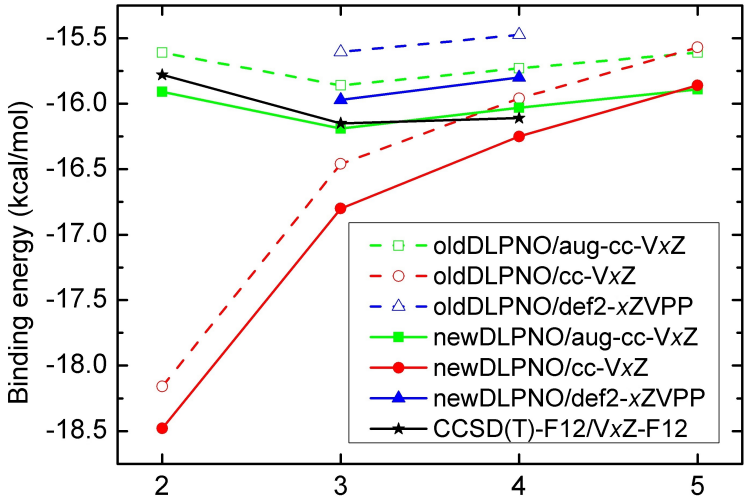

Figure 11: Binding energies (kcal/mol) for sulfuric acid–ammonia complex using RI-CCSD(T)-F12 and DLPNO–CCSD(T) methods as a function of basis set cardinal number x. Old refers to the Orca3 version of DLPNO with default criteria, and new in turn to the Orca4 DLPNO with TightPNO criteria.

The new linear scaling DLPNO method as well as the three default thresholds have been implemented in Orca version 4, and were only released for use in 2017.[303] Therefore, in the studies of this thesis **Papers I–VI** as well as **Papers VIII**, **X**, and **XI**, the old DLPNO–CCSD(T) method was utilized.

3.2.9 Semi-empirical Quantum Chemistry Methods

Semi-empirical methods are based on the Hartree–Fock formalism, but several approximations have been introduced and some parameters have been obtained from empirical data.[304] Empirical parameters allow inclusion of some electron correlation into the methods. In contrast to the HF approach, semi-empirical methods are fitted to a set of empirical parameters and the two-electron part of the Hamiltonian is not explicitly included. Because some of the two-electron integrals are approximated, semi-empirical calculations are much faster than *ab initio* calculations. The central assumption of semi-empirical models is the zero differential overlap (ZDO) approximation, which neglects all products of basis functions that depend on the same electron coordinates when located on different atoms.[305] The error introduced using semi-empirical methods is compensated through the use of parameters determined comparing calculations with experimental data.

Semi-empirical methods can be grouped as complete neglect of differential overlap (CNDO),[306,307] intermediate neglect of differential overlap (INDO),[308,309] and neglect of diatomic differential overlap (NDDO) approaches.[310,311] **Papers III–VI**, **VIII**, and **XI** utilize the parametric method number 6 (PM6) for initial optimization of a large set cluster structures. PM6 is an NDDO type semi-empirical method, which is parametrised for 70 elements. Compared to previous NDDO models, the main improvement of PM6 is that the prediction of the energies and geometries involves in hydrogen bonding. The detailed parametrisation can be found from Ref. 181.

3.3 Density Functional Theory

Density functional theory (DFT) is based on the theorems of Hohenberg and Kohn,[312] which state that the ground-state electron density uniquely determines the electronic energy according to the variational principle. The electron density has the same number of variables independent of the system size, since each spin density depends only on three spatial coordinates. In modern Kohn-Sham DFT the energy is written as a functional of the electron density, and the effective external potential is generated from a fictive non-interacting reference system.[313] All parts of the Hamiltonian cannot be formulated in this way, and in addition the parameters which define a given functional have to be introduced. Therefore, DFT is not systemically improvable.[314]

The general form of the Kohn–Sham DFT energy functional is

$$E_{\mathrm{DFT}}[\rho(\mathbf{r})] = T_{\mathrm{e}}[\rho(\mathbf{r})] + V_{\mathrm{ne}}[\rho(\mathbf{r})] + V_{\mathrm{ee}}[\rho(\mathbf{r})] + E_{\mathrm{XC}}[\rho(\mathbf{r})], \tag{31}$$

where $T_{\mathrm{e}}[\rho(\mathbf{r})]$ is the kinetic energy of non-interacting electrons, $V_{\mathrm{ne}}[\rho(\mathbf{r})]$ is the attractive interaction between nuclei and electrons, $V_{\mathrm{ee}}[\rho(\mathbf{r})]$ is the repulsive interaction between electrons, and $E_{\mathrm{XC}}[\rho(\mathbf{r})]$ is the exchange–correlation term. In other words, the Schrödinger equation is reformulated in terms of the electron density, and exchange and correlation interactions are included via an exchange–correlation functional. Unfortunately, the exact exchange–correlation functional — which connects the electron density to the kinetic energy and the electron-interaction energies — is unknown.[313] Otherwise the obtained solution would be exact. Hence the main task in Kohn–Sham DFT is to derive approximations to the exchange–correlation energy functional.

3.3.1 Density Functionals

The accuracy of DFT methods is dependent on the exchange–correlation functional. Since it is not known, various approximate functionals have been developed to calculate different molecular properties.[315] There is no straightforward way to say whether one functional is better than another. On the basis of the analysis of Perdew, exchange–correlation functionals can be divided into five levels along Jacob's ladder so that successive levels correspond to better approximations that bring us closer to the heaven of chemical accuracy.[316]

The first level on Jacob's ladder is called the local density approximation (LDA), which approximates the exchange–correlation energy density at a given position as a function of the electron density at that same local position.[317] Since the density is assumed to be a slowly varying function in LDA, but the electron density is typically rather far from spatially uniform in a real chemical system, the LDA is not useful for molecular systems. However, it can be used to model for example metal surfaces, where the electron density of the system varies gradually.[318]

The second level is the generalized gradient approximation (GGA),[319] for which the electron density approximation also depends on the gradient of the density at that given position.[320] Since the exchange and correlation energies depend not only on the electron density but also on the derivatives of the density, GGAs have shown to be a significant improvement over LDAs.[321] At the third level comes the meta-GGA,[322] for which the exchange and correlation functionals depend on higher order derivatives of the electron density or on the local kinetic energy density of the Kohn–Sham orbitals. The computational cost of a meta-GGA is comparable to that of a GGA, and the meta-GGA is typically (but not always) a bit more accurate than the GGA.[323]

The fourth and fifth levels of Jacob's ladders are formed by combining the GGA or meta-GGA functional with an amount of exact Hartree–Fock exchange or MP2 correlation energy, called hybrid and double-hybrid functionals, respectively. The most popular hybrid functional is B3LYP, which combines the Becke-3-parameter exchange functional with Lee–Yang–Parr correlation functional.[324–326] An example of a double hybrid functional is B2PLYP, which contains B88 exchange, two parameters that were fitted and perturbative mixture of MP2 and LYP.[327] In addition, there are other methods to include the virtual orbitals, for instance the random phase approximation. Inclusion of the virtual space is expected to yield large improvement on dispersion interaction effects, which are a significant problem for almost all common functionals.[318]

Dispersion Corrections

Dispersion interactions depend on electron correlation, but most of the functionals neglect the long-range dispersion and only local contributions to the electron correlation are included. Typically, DFT functionals do not model correlation outside the Fermi hole (which arises from the fact that electrons have a low probability of being found close to each other), and thereby neglect long-range dispersion effects.[328] Notably, the quality of modelling dispersion effects is highly functional dependent. For instance, when considering equilibrium distances for van der Waals complexes, some density functionals such as PW91 provide at least qualitatively correct interaction potentials,[329] whereas some other functionals like B3LYP and BLYP predict purely repulsive potentials.[318] Therefore, functionals that are able to model dispersion interactions are extensively developed. For example, density functionals such as M06-2X have been parametrized to systems governed by dispersion interactions.[330] Implicit inclusion of dispersion has shown some success for describing weakly-bound complexes.[331] Both PW91 and M06-2X have been shown to perform well for atmospheric molecular clusters (see **Papers I–III**).[332,333] **Paper IX** shows that the PW91 functional yields good transition state structures, and it can be utilized for geometries and frequencies when studying organic reaction mechanisms. It should be mentioned that M06-2X as well as other Minnesota functionals converge remarkably slowly towards the complete basis set limit, especially when calculating intermolecular interaction energies (see **Paper II**).[334]

Most of the current dispersion-corrected approaches includes the dispersion interactions as an external correction to the density functional. Grimme has suggested the DFT-D approach,[335] which treats dispersion as an additional empirical term E_{disp} to the DFT energy E_{DFT} as

$$E_{\text{DFT-D}} = E_{\text{DFT}} + E_{\text{disp}}. \tag{32}$$

Since the dispersion correction is an add-on term, it does not directly alter the wave function, electron density, or any other molecular property.[336] However, geometry optimizations with dispersion corrections will lead to a different geometry than without because the dispersion correction contributes to the forces acting on the atoms.[337] Mardirossian and Head-Gordon have developed a range-separated hybrid-GGA ωB97X-D functional, which contains ten parameters and nonlocal correlation effects.[338] It has been demonstrated to predict good structures and thermochemical parameters for non-covalent molecular clusters (see **Papers I–III**).[332,339] Also other approaches to include dispersion corrections have been developed, for example, non-local van der Waals density functionals (vdW-DFs)[340–342] and atom-centered one-electron potentials (1ePOT).[343–345]

3.4 Thermochemical and Vibrational Analysis

The electronic structure calculations discussed above only yield ground state electronic energies at the temperature T=0 K. In order to obtain free energies, thermochemical parameters must be taken into account. Molecular systems have translational, vibrational, and rotational degrees of freedom. A usual way to proceed is to first assume that the different energetic contributions are uncoupled, meaning that the different degrees of freedom can be separated from each other.[186] Then the total energy (ϵ_{tot}) can be written as a sum of translational (ϵ_{tr}), rotational (ϵ_{rot}), vibrational (ϵ_{vib}), and electronic (ϵ_{el}) energies as

$$\epsilon_{\text{tot}} = \epsilon_{\text{el}} + \epsilon_{\text{tr}} + \epsilon_{\text{vib}} + \epsilon_{\text{rot}}, \tag{33}$$

and the partition function for the system q_{tot} can be expressed as a product of four components

$$q_{\text{tot}} = q_{\text{el}} q_{\text{tr}} q_{\text{vib}} q_{\text{rot}}. \tag{34}$$

The Gibbs free energy G_{tot} is dependent on the total enthalpy H_{tot} and entropy S_{tot} as

$$G_{\text{tot}} = H_{\text{tot}} - T S_{\text{tot}}. \tag{35}$$

For an uncoupled system, the total enthalpy and entropy can be expressed as a sum of the four contributions

$$H_{\text{tot}} = H_{\text{el}} + H_{\text{tr}} + H_{\text{vib}} + H_{\text{rot}} \tag{36}$$

$$S_{\text{tot}} = S_{\text{el}} + S_{\text{tr}} + S_{\text{vib}} + S_{\text{rot}}. \tag{37}$$

Each component of enthalpy H_X is calculated from partition functions q_X as following

$$H_X = k_B T^2 \left(\frac{\partial \ln q_X}{\partial T} \right)_V + k_B T V \left(\frac{\partial \ln q_X}{\partial V} \right)_T \tag{38}$$

and the entropy components S_X are calculated as

$$S_X = k_B T \left(\frac{\partial \ln q_X}{\partial T} \right)_V + k_B \ln \left(q_X \right). \tag{39}$$

The gap of electronic energy between the ground state and the lowest excited state is normally large enough that the excited states can be assumed to be unpopulated at chemically interesting temperatures.[199] Thus the electronic contribution is commonly taken to be the ground state energy of the global minimum energy structure, and the electronic partition function is then

$$q_{\text{el}} = g_0 e^{\frac{-E_0}{k_B T}}, \tag{40}$$

where E_0 and g_0 denote the ground-state electronic energy and degeneracy, respectively. For non-degenerate singlet ground-state wave functions $g_0 = 1$.

Usually, the translational motion of the systems is taken to be that of ideal gas particles, and the translational partition function can be written as

$$q_{\text{tr}} = \left(\frac{2\pi M k_{\text{B}} T}{h^2}\right)^{\frac{3}{2}} \left(\frac{k_{\text{B}} T}{p}\right), \tag{41}$$

where h is Planck constant, p is the partial pressure, and M is the mass. The ideal gas volume is defined as $V = \frac{k_{\text{B}} T}{p}$.

The rotational movement of a system is often assumed to be rigid and independent of the rotational and vibrational quantum numbers.[217] Thus the parameters needed for the calculation are the rotational symmetry number s of the system, which is typically equal to 1 for larger structures, and the moments of inertia about the principal axes of inertia I_n. The rotational partition function can be approximated using classical mechanics as

$$q_{\text{rot}} = \frac{\sqrt{\pi}}{s} \left(\frac{8\pi^2 k_{\text{B}} T}{h^2}\right)^{\frac{3}{2}} \sqrt{I_1 I_2 I_3}. \tag{42}$$

Using the harmonic oscillator approximation, the vibrational degrees of freedom are uncoupled in a normal coordinate system, and the vibrational partition function for non-linear systems can be written over $3N - 6$ vibrational modes as

$$q_{\text{vib}} = \prod_{i=1}^{3N-6} \frac{e^{\frac{-h\nu_i}{2k_{\text{B}}T}}}{1 - e^{\frac{-h\nu_i}{k_{\text{B}}T}}}, \tag{43}$$

where N is the number of atoms and ν_i is the frequency corresponding to normal mode i. The computation of vibrational partition function is challenging since it requires a vibrational analysis, *i.e.*, the second derivatives of the electronic energy with respect to $3N$ nuclear coordinates must be calculated.[188] It should be emphasized that vibrational analysis is valid only when the first derivatives of the energy with respect to displacement of the atoms are zero. The gradient g holds the first partial derivatives of the potential U with respect to displacement of the atoms in cartesian coordinates q_n as

$$g = \begin{bmatrix} \frac{\partial U}{\partial q_1} \\ \frac{\partial U}{\partial q_2} \\ \vdots \\ \frac{\partial U}{\partial q_n} \end{bmatrix} = 0. \tag{44}$$

In practise this means that the geometry used for vibrational analysis must be optimized at the same level of theory as the second derivatives were generated with. The Hessian H is a $3N \times 3N$ matrix and it can be expressed as the second partial derivatives of the potential U with respect to displacement of the atoms

$$H = \begin{bmatrix} \frac{\partial^2 U}{\partial q_1^2} & \cdots & \frac{\partial^2 U}{\partial q_1 \partial q_n} \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 U}{\partial q_n \partial q_1} & \cdots & \frac{\partial^2 U}{\partial q_n^2} \end{bmatrix}. \tag{45}$$

Diagonalization of the mass weighted Hessian matrix yields the $3N$ eigenvectors of the system, namely the *normal modes*. Three of the normal modes are related to the translational motion and three to the rotational motion (for non-linear systems), hence there are a total of $3N - 6$ vibrational modes.[186] In theory, the frequencies for rotation and translation modes should be zero, but due to the numerical disturbance they remain close to zero in practical calculations. If all the vibrational frequencies are positive the optimized geometry corresponds to the minimum energy structure. When one negative frequency is obtained the structure corresponds to a transition state, which means that the energy is maximized along one Hessian eigenmode and minimized along the remaining $3N - 7$ eigenmodes.

Within the rigid rotor-harmonic oscillator (RRHO) approximation, the molecular systems are assumed to behave as equilibrated ideal gas particles and the effect of temperature is included by the molecular structures vibrating harmonically about their equilibrium geometries and rotating rigidly as a single entity.[188] Unfortunately, real molecular systems are not rigidly rotating harmonic oscillators, and the RRHO approximation may lead to large uncertainties and even non-physical thermal contributions to the Gibbs free energy (see **Paper II**).[92,346] Therefore, numerous approaches have been developed to at least partially correct for the worst errors of the RRHO approximation.

Anharmonic calculations for large molecules are difficult due to the nonseparability of the Hamiltonian. The simplest way to account for vibrational anharmonicity is to derive scaling factors for small systems and simply apply them for larger systems by multiplying the harmonic frequencies by the scaling factor in an *ad hoc* fashion in the standard formulae.[347–350] This approach is used in **Paper II** for large acid–base clusters. It should be noted that for instance intermolecular bonds are normally more anharmonic than intramolecular bonds, but by applying scaling factors different types of vibrations are treated similarly.

For medium-sized systems, one possible way to calculate anharmonic vibrational spectra is to apply the *ab initio* vibrational self-consistent field (VSCF) approximation.[351] The

calculations can be performed from the first principles, and thus fitting of PES or using empirical parameters are not needed. The vibrational problem is treated in a similar manner as the HF approach so that in any given vibrational state of the system each vibrational mode is described by the averaged potential of all other modes. The VSCF potential requires multidimensional grid-point calculations, which are subsequently used for the numerical solution of the one-dimensional VSCF equations. The determination of the VSCF Hamiltonian requires on the order of $6N$ single-point energy calculations. Since all the VSCF equations are solved numerically until the convergence is reached, the convergence might cause problems for some systems such as non-covalently bound molecular clusters. The VSCF results can be improved by adding correlation between the different modes, for example, using the second order perturbation theory (VSCF-PT2)[352] as in **Paper VII**.

Another option is to treat the anharmonicity as a perturbation to the RRHO system and calculate the anharmonic oscillator energy levels using perturbation theory. The method is called vibrational second order perturbation theory (VPT2).[353] In VPT2 the anharmonic corrections are calculated from third and fourth order derivatives of the PES along the normal mode coordinates. The cubic and semi-diagonal quartic force constants are calculated by finite differentiation of the Hessian. If analytical second derivatives are available, then the required third and fourth derivatives can be computed easily using finite differentiation. The cost of VPT2 level is in the order of $6N$ times that of a single harmonic vibrational calculation. The VPT2 method yields only frequencies and not intensities, and thus it is not commonly used for spectroscopy studies. Another drawback of VPT2 is that it is subject to the problem of near degeneracies.

In 2011, Temelso *et al.* investigated the role of anharmonicity in hydrogen-bonded water clusters.[354] They determined scaling factors by comparing harmonic vibrational frequencies with VPT2 anharmonic fundamentals of 16 water clusters. They found that the intermolecular modes are substantially more anharmonic than intramolecular bending and stretching modes, which demonstrates the importance of separating the modes into different classes. Also the disparity in derived scaling factors between water clusters and covalently-bound molecules highlights the need to apply different scaling factors for hydrogen-bonded systems. Temelso *et al.* showed that the anharmonic effects lower the Gibbs free formation energy substantially, however, the energetic ordering of different isomers remain the same.[355] For large acid–base clusters, **Paper II** shows that vibrational anharmonicity affects a maximum lowering of 2 kcal/mol in the thermal contribution to the Gibbs free energy. However, the error arising from anharmonic frequencies is significantly smaller than the error due to the high-amplitude

modes.

In addition to anharmonicity, internal rotation effects are inherent in most large systems, for instance molecular clusters, for which molecules or some part of the cluster can pseudorotate freely.[356] Then the vibrational potential has multiple shallow minima and the motion is better described as an internal rotation or a pseudorotation.[357] However, the identification of the internal rotations is difficult since the low-lying frequency modes might include both internal rotations and large amplitude collective bending motions of atoms, and in addition, some of them might be a mixture of both.[358] Moreover, there is no simple analytic solution for hindered rotor problems and special approximations must be used.[359] Different approaches to correct partition functions, when internal rotations are identified, are suggested in the literature.[356,358,360–363] However, these methods often require the use of internal coordinates, which have not been extended to non-covalently bound systems.[358,364] To avoid these issues, Grimme has suggested so-called quasi-harmonic approximation (QHA),[364] where the low-lying frequencies are treated as a free rotor. This approach is utilized in **Paper II** to correct the failure of the RRHO approximation in weakly-bound molecular clusters.

The vibrational entropy of a harmonic oscillator with frequency ω is

$$S_{\mathrm{V}} = R\left[\frac{h\omega}{k_{\mathrm{B}}\left(e^{h\omega/k_{\mathrm{B}}T}-1\right)} - \ln\left(1-e^{-h\omega/k_{\mathrm{B}}T}\right)\right], \tag{46}$$

where R is the gas constant. The second term of Equation (46) approaches asymptotically infinity for $\omega \to 0$ yielding an unphysical entropy at low frequency values. In QHA, the contribution of low-lying frequencies to the entropy is replaced by a corresponding rotational entropy as

$$S_{\mathrm{R}} = R\left\{\frac{1}{2} + \ln\left[\left(\frac{8\pi^3\mu' kT}{h^2}\right)^{1/2}\right]\right\}, \tag{47}$$

where μ' is an effective moment of inertia, calculated from the moment of inertia μ for a free-rotor and the average moment of inertia B_{av} as Equation (48). An effective moment of inertia is restricted to reasonable values by using $B_{av} = 10^{-44}$ kg m^2 as a limiting value for very small ω. An effective moment of inertia can be written as

$$\mu' = \frac{\mu B_{av}}{\mu + B_{av}}. \tag{48}$$

A weighting function is used to interpolate between harmonic vibrational entropy S_{V} for $\omega >> \omega_0$ and pure rotational entropy S_{R} for small ω close to the cut-off frequency ω_0 as follows

$$S = w(\omega)S_{\mathrm{V}} + [1 - w(\omega)]S_{\mathrm{R}}, \tag{49}$$

where $w(\omega)$ is the Head-Gordon damping function (with $\alpha = 4$)[365]

$$w(\omega) = \frac{1}{1 + \left(\frac{\omega_0}{\omega}\right)^{\alpha}}. \tag{50}$$

Now the only parameter which has to be fixed is the cut-off value ω_0. **Paper II** applies a cut-off frequency of 100 cm^{-1}, as proposed in the literature.[365,366] In principle, the QHA partition function should be close to the ideal hindered rotor partition function with a reasonable cut-off frequency. The QHA method does not require any extra computational power, and therefore, it can be applied to large molecular clusters. Funes-Ardois and Paton have implemented QHA in the Python script GoodVibes.py,[367] and it is also available in Orca version 4.

3.5 Gibbs Free Formation Energy

When all the partition functions are known, the Gibbs free formation energy for the cluster AB formed from monomers A and B can be calculated as

$$\Delta G = \Delta E - k_{\mathrm{B}} T \ln \left[\frac{q_{\mathrm{vib}}(AB) q_{\mathrm{rot}}(AB) q_{\mathrm{tr}}(AB)}{q_{\mathrm{vib}}(A) q_{\mathrm{vib}}(B) q_{\mathrm{rot}}(A) q_{\mathrm{rot}}(B) q_{\mathrm{tr}}(A) q_{\mathrm{tr}}(B)} \right], \tag{51}$$

where ΔE corresponds to the ground-state electronic energy difference between the cluster and its monomers. The largest contribution to the Gibbs free energy comes from the ground-state electronic energy, and one can visualise that the procedure described above to be a kind of extrapolation scheme from the zero-temperature to desired temperature.[186] Marking the second term of Equation (51) as the thermal contribution to the Gibbs free energy ($\Delta G_{\mathrm{therm}}$), the Gibbs free formation energy can be simply written as

$$\Delta G = \Delta E + \Delta G_{\mathrm{therm}}. \tag{52}$$

3.5.1 Towards a Cost-Effective and Robust Approach

The aim of most quantum chemical studies is to obtain the best possible accuracy of the properties of interest with minimal computational costs. Since the accuracy as well as required resources are strongly dependent on the combination of the method and basis set, finding a suitable level of theory for a specific problem is of utmost importance (see **Papers I** and **II**).[239,332] It should be kept in mind that the best possible result obtained with a given level of theory does not mean the correct result in absolute sense. From a chemical point of view

the accuracy of absolute energy is not important whereas the relative energies, *i.e.* the energy differences along to the reaction coordinate, are the values of interest.

Often experience and literature offer a good starting point for finding a proper quantum chemical level which leads to trustworthy results. The comparison with highly accurate benchmark calculations or with reliable measurement data is further needed to confirm the applicability for a specific problem. Geometry optimization and vibrational frequencies are less sensitive to the level of theory used in the calculations than the energies themselves (see **Papers I–III**).[332,333] As discussed in the previous section, the second partial derivatives of the energy must be computed in order to get the thermochemical properties, meaning that it is computationally the most expensive part. Density functional theory often yields reliable molecular geometries and vibrational frequencies and is therefore commonly used. In addition, relatively small basis sets are usually suitable for DFT methods since they converge relatively fast. This means that a larger basis set would only have a negligible effect on these quantities but a large effect on computational effort (see **Papers II** and **VI**).[332,368] In contrast, the electronic energies are very sensitive to the level of theory, and correlated wave function methods are usually needed to obtain energies within the chemical accuracy, especially for non-covalently bound molecular clusters (see **Paper I**).[339,369] Unfortunately, correlated wave function methods converge slowly with respect to the basis set size, and a large basis set is necessary for accurate energy calculations (see **Papers VI** and **VIII**). Due to these reasons, it is common practice to optimize the geometry and to calculate thermochemical properties at a lower level of theory, and then perform the electronic energy corrections using more sophisticated correlated methods (see **Papers I–VI** and **VIII–XI**).

3.5.2 Review of Computational Approaches used in Atmospheric Clustering Studies

Nadykto *et al.* has utilized the PW91/6-311++G(3df,3pf) level of theory to calculate Gibbs free energies in atmospheric nucleation studies.[89,108,370–373] Gibbs free energies, especially for large clusters, are not experimentally directly measurable quantities, and thus the comparison between computational and experimental results has to be made through cluster kinetics.[143] Kjærgaard *et al.* has determined experimentally Gibbs free energies for small two-component molecular clusters (the complexes of methanol or ethanol with dimethylamine).[374] However, in practice the measurable quantity for observing particle formation processes is the cluster concentration.[60] From a theoretical point of view this means that extra approximations are needed to model large sets of interacting clusters. Therefore, other error sources might become

important. Answering questions such as
"How well can kinetic gas theory describe molecular cluster collisions?"
"What do experimental instruments actually measure?"
"What is the sensitivity of an instrument to detect clusters with low concentrations?"
"How much does the composition of neutral cluster change when charging it before detecting?"
is an extremely difficult task. Hence, one should be cautious when making conclusions from experimental and theoretical comparisons. Nadykto *et al.* has showed that nucleation rates calculated using the PW91/6-311++G(3df,3pf) level give a good correspondence with experimental nucleation rates.[372,375,376] **Paper II** shows that the PW91 functional with the 6-311++G(3df,3pd) basis set gives a mean absolute error of 0.9 kcal/mol in the binding energy compared to a large aug-cc-pV5Z basis set, indicating that it has not reached the complete basis set limit. In addition, **Paper III** shows that PW91/6-311++G(3df,3pd) yields a difference of 2.5 kcal/mol in the binding energies compared to the CCSD(T)-F12a/VDZ-F12 level. Therefore, there is reason to believe that the good correspondence is originating from cancellation of errors — and thus it would not be systematic — or it is also possible that experimental and theoretical findings do not describe the exactly same quantity.

Kurtén *et al.* has utilized the RI-CC2 method for single point energies to improve the quality of the binding energies.[84,238] RI-CC2 is an approximate coupled cluster singles and doubles method with a resolution of the identity approximation.[237] In 2008, Ortega *et al.* has introduced a B3RICC2 multi-step approach,[239] where the geometry is optimized and frequencies are calculated at the B3LYP/CBSB7 level and single point energies are calculated using the RI-CC2 method with an aug-cc-pV(T+d)Z basis set. The B3RICC2 approach has commonly been used in atmospheric new-particle formation studies.[64,151,377–380] The idea is in principle well founded, since the electronic energy corrections eliminate random errors in DFT binding energies. However, as Hättig has pointed out, the CC2 method is developed for excited states,[236] and accordingly overestimates the correlation effect of ground states, *i.e.*, binding energies tend to be too negative. In addition, it has been shown that CC2 energies are usually not more accurate than MP2 energies for ground states, and that actually the overbinding of CC2 might be even higher than that of MP2.[380,381] RI-CC2 is approximately 10–20 times more expensive than MP2 calculations due to an iterative solution of the cluster equations. Therefore, it is difficult to justify the usage of CC2 even over the MP2 method. Furthermore, the CBSB7 basis set corresponds to Pople's triple-zeta basis, which contains two additional d polarization functions on second rows atoms, one d function on first row atoms and a p function on hydrogen atoms, but it includes no diffuse functions (6-311G(2d,d,p)).[382] **Paper II** shows the importance of including diffuse functions. The B3LYP functional is also

well-known to give inaccurate binding energies and it is not recommended for non-covalently bound systems where dispersion interactions have a significant role (see **Paper I**).[369,383]

Table 1 compares Gibbs free formation energies for sulfuric acid–ammonia and sulfuric acid–dimethylamine clusters calculated with different levels of theory. The highest theory level, RI-CCSD(T)-F12/VQZ-F12//M06-2X/6-31++G**, yields Gibbs free formation energies of -5.3 and -11.5 kcal/mol for $(H_2SO_4)(NH_3)$ and $(H_2SO_4)(NH(CH_3)_2)$, respectively. The RI-CC2 energy corrections yield a significant overbinding (always more than 1 kcal/mol). In the case of $(H_2SO_4)(NH(CH_3)_2)$ when using B3RICC2, the overbinding is up to 3.9 kcal/mol. The PW91/6-311++G(3df,3pd) level performs well for $(H_2SO_4)(NH(CH_3)_2)$ but it overbinds 2.5 kcal/mol for $(H_2SO_4)(NH_3)$. This implies that the good correspondence for $(H_2SO_4)(NH(CH_3)_2)$ is due to a lucky cancellation of errors, and thus even the sign of the PW91 error is not systematic. The DLPNO–CCSD(T)//DFT level (see Section 4.1) systematically underbinds, with an error of less than 1 kcal/mol for both systems in Gibbs free formation energies.

Table 1: Gibbs free formation energies (kcal/mol) for sulfuric acid–ammonia and sulfuric acid–dimethylamine complexes at 298.15 K and 1 atm.

Method, Ref	$(H_2SO_4)(NH_3)$	$(H_2SO_4)(NH(CH_3)_2)$
B3RICC2, 239	-6.4	-15.4
RI-CC2/aV(T+d)Z//RI-MP2/aV(D+d)Z, 84	-6.6	-13.7
PW91/6-311++G(3df,3pd), 89	-7.8	-11.4
DLPNO–CCSD(T)//DFT, **Paper III**	-4.6	-10.6
RI-CCSD(T)-F12/VQZ-F12//M06-2X/6-31++G**	-5.3	-11.5

4 Results and Discussion

This chapter presents the main results obtained in **Papers I–VI**. The author's contribution in the research articles of this thesis is specified. The author is solely responsible for the introductory part of the thesis.

4.1 Computational Methods Assessment

When studying the molecular-level formation mechanisms and stabilities of atmospheric molecular clusters, the equilibrium structures as well as vibrational frequencies for all participating monomers and clusters are needed. Thermochemical parameters can be calculated once the vibrational frequencies are known. This means that a lot of computational power is needed, and the most reasonable method of choice is thereby DFT. Single point energy calculations require much less computational power than the geometry optimization and frequency calculations, and once the structure is optimized, electronic energy corrections can be calculated with a higher level method. All geometry optimizations and frequency calculations have been run using Gaussian09 revisions B.01 and D.01.[384] All DLPNO–CCSD(T) single point energies have been computed using Orca version 3.0.3.[293] For CCSD(T)-F12 calculations, both Molpro (versions 2012.1 and 2015.1)[385] and Orca have been used.

4.1.1 Sensitivity and Accuracy of Density Functional Theory

In order to know which functional and basis set combinations are suitable for modelling non-covalently bound cluster formation, we have tested the performance of different levels of theory for predicting structures, thermochemistry, and binding energies (**Papers I–III**). In **Paper I**, we have chosen six small cluster formation reactions to represent some of the key interactions in atmospheric molecular clusters. We have investigated the variation in the Gibbs free energy using 11 different functionals with the 6-311++G(3df,3pd) basis set[386,387] as well as MP2 with aug-cc-pV(D+d)Z and aug-cc-pV(T+d)Z basis sets.[388] We found a large scatter in the binding energies depending on which DFT functional is utilized, with variations up to 4.0 kcal/mol between PW91 and CAM-B3LYP-D. However, only a little variation was observed between different functionals in the thermal contributions to the Gibbs free energy, with the largest difference of 1.1 kcal/mol between M11 and B3LYP-D. The use of coupled cluster electronic energy corrections on top of the DFT or MP2 geometries significantly reduces the scatter in the binding energies, with the largest variation being reduced to 0.6 kcal/mol. This indicates that all tested functionals can be used to obtain geometries and frequencies, but electronic energy corrections are needed to reduce the scatter in binding energies.

In **Paper III** we have studied the sensitivity and accuracy of M06-2X, PW91, and ωB97X-D functionals with 6-31++G(d,p) and 6-311++G(3df,3pd) basis sets[389] compared to the DLPNO–

CCSD(T)/def2-QZVPP and the CCSD(T)-F12a/VDZ-F12 levels of theory. Table 2 shows the binding energies and thermal contributions with different levels of theory for a small diperoxy acid–sulfuric acid complex $(C_3H_4O_6)(H_2SO_4)$. Figure 12 presents the structure of the $C_3H_4O_6$ diperoxy acid molecule, which is used as a model compound in benchmark studies.

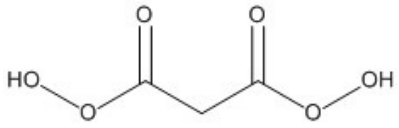

Figure 12: The structural formula of $C_3H_4O_6$ compound.

Table 2: Thermal contributions to the Gibbs free energy ΔG_{therm} (kcal/mol) using DFT at 298.15 K and 1 atm and binding energies ΔE at DFT, DLPNO–CCSD(T)/def2-QZVPP, and CCSD(T)-F12a/VDZ-F12 levels of theory for the $(C_3H_4O_6)(H_2SO_4)$ cluster.

	ΔG_{therm}	ΔE_{DFT}	ΔE_{DLPNO}	ΔE_{F12}
6-31++G(d,p)				
M06-2X	14.9	-21.7	-15.5	-16.9
PW91	14.2	-15.6	-15.0	-17.6
ωB97X-D	13.4	-19.5	-16.2	-17.5
6-311++G(3df,3pd)				
M06-2X	15.4	-20.4	-16.1	-17.4
PW91	14.4	-14.8	-15.7	-17.3
ωB97X-D	13.3	-18.4	-16.3	-17.5

We showed in **Paper III** that different functionals yield highly varying binding energies for $(C_3H_4O_6)(H_2SO_4)$ with a maximum variation of 6.9 kcal/mol. We found that the thermal contribution to the Gibbs free energy varies much less, with values from 13.3 to 15.4 kcal/mol. A negligible change in the thermal contribution was detected when reducing the basis set from 6-311++G(3df,3pd) to 6-31++G(d,p). We observed only a small variation in the DLPNO–CCSD(T)/def2-QZVPP binding energies with a maximum of 1.3 kcal/mol and in the CCSD(T)-F12a/VDZ-F12 binding energies with a maximum of 0.7 kcal/mol. The DLPNO binding energies underbind compared to canonical coupled cluster with an average underbinding of 1.6 kcal/mol. The ratio between the coupled cluster results is found to vary with an average value of 1.10. This implies that the DLPNO–CCSD(T)/def2-QZVPP level of theory can be used as a lower bound for the binding energies.

Paper II examines the basis set convergence of M06-2X, PW91, and ωB97X-D functionals by comparing different basis sets with a large aug-cc-pV5Z basis to make sure that the DFT/basis combination yields results close to the basis set limit. Figure 13 presents the accuracy of binding energies compared to the aug-cc-pV5Z basis set (red bars) and the relative time compared to the smallest 6-31G* basis set (blue squares). Including diffuse functions is known to be important for a correct description of loosely bound electrons, however, **Paper II** shows that in most cases partially augmented[390] basis sets yield as accurate results as the fully augmented basis sets with significant gain in computational efficiency. Double-zeta basis sets with some diffuse functions are sufficient for geometry optimizations and thermochemical parameters.[368] For binding energies, at least augmented triple-zeta basis sets are needed to reach the complete basis set limit.

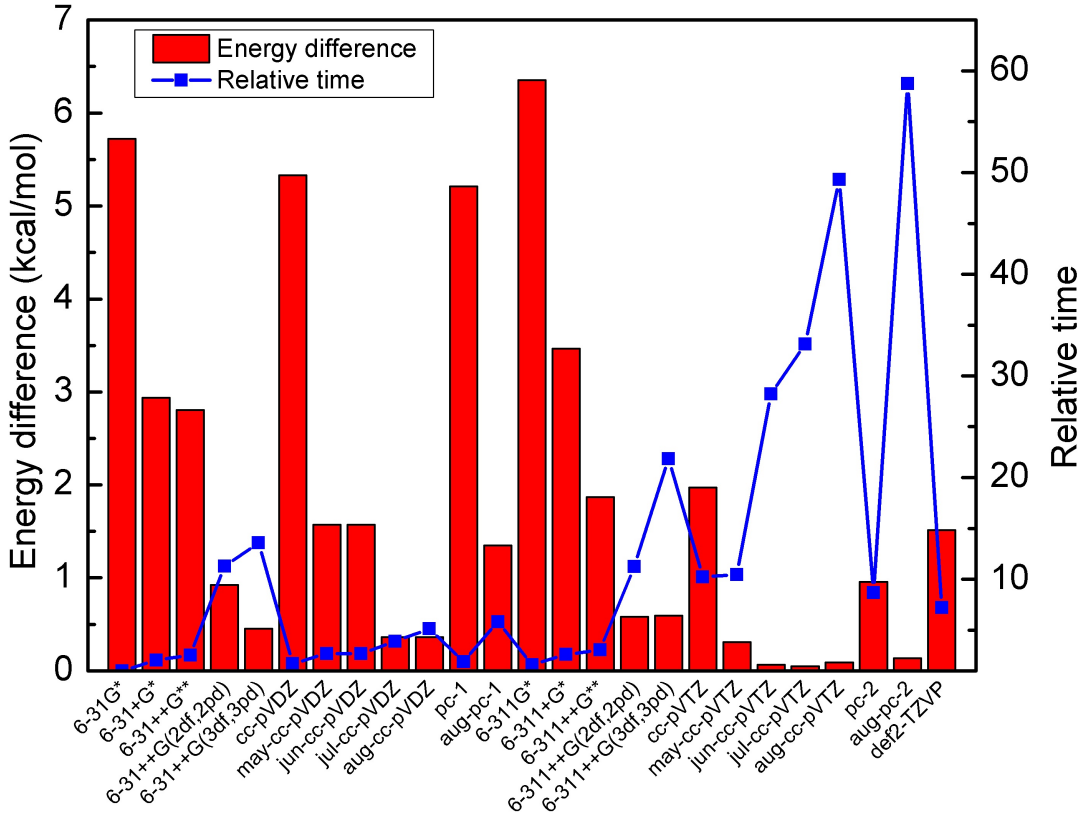

Figure 13: Relative time (blue squares) and accuracy of binding energies (red bars) for the $(H_2SO_4)(NH_3)$ cluster calculated using the PW91 functional with double and triple-zeta basis sets. Figure modified from **Paper II**.

The tested functionals, M06-2X, PW91, and ωB97X-D, appear to perform relatively similar for structures and thermal contributions over the studied cluster formation reactions. However, in some cases a single density functional might vary up to a few kcal/mol. To compensate these outliers, we recommend to average over multiple functionals by utilizing more than one functional and calculating the average value. In **Papers III–VI** we have used M06-2X, PW91, and ωB97X-D functionals with a 6-31++G** basis set for predicting structures and thermochemistry. In binding energies, the variation between different functionals remains large even with a large basis set. The problem with DFT is that there is no straightforward way to say which functional performs best. Therefore, we recommend to calculate single point energies using a high-level correlation method on top of the DFT structures to eliminate the large variation between functionals.

Paper II compares errors arising from vibrational anharmonicity and low-lying vibrations in rigid rotor-harmonic oscillator approximation. We derived anharmonic scale factors using vibrational second order perturbation theory (VPT2) and utilized the quasi-harmonic approximation as suggested by Grimme.[364] We showed that when the low-lying frequencies ($\leq$100 cm^{-1}) are treated separately, the errors arising from the anharmonicity of the remaining frequencies are small regardless of system size. However, the quasi-harmonic correction for low-lying frequencies is a few percent of the total harmonic thermal contributions, when using a cut-off value of 100 cm^{-1}, which is much larger than the error arising from vibrational anharmonicity. We suggested that the study of the accuracy of the quasi-harmonic approximation should be extended to a wider test set since low-lying frequencies seem to be significant error source for large molecular clusters.

In order to investigate whether the findings for neutral clusters are still valid for anionic clusters, in **Paper VI** we have computed thermal contributions to the Gibbs free energy for negatively charged molecular clusters using the M06-2X functional with several double and triple-zeta basis sets. We have taken a test set of 4 anionic cluster formation reactions and calculated the mean absolute errors (MAEs) relative to the largest aug-cc-pVTZ basis set, and the relative computational times relative to the smallest cc-pVDZ basis set (see Figure 14). This study shows that a small 6-31++G** basis set is indeed sufficient in order to obtain the thermochemical parameters also for anionic clusters.

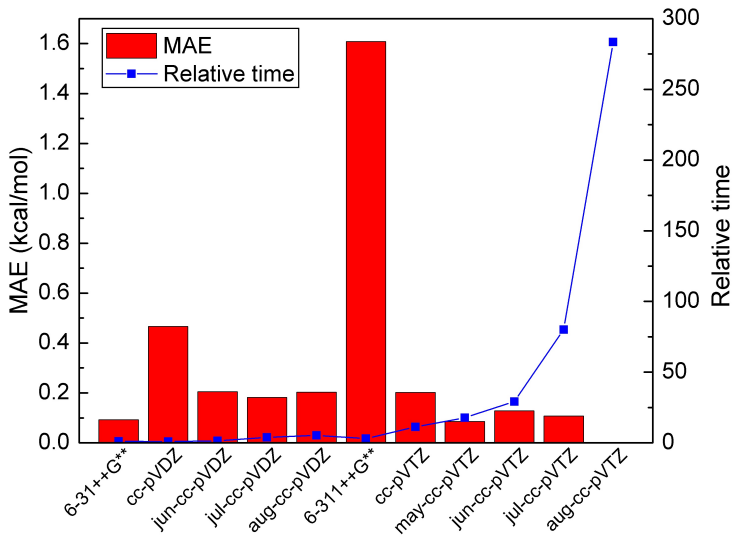

Figure 14: Mean absolute errors in the thermal contribution (kcal/mol) relative to the aug-cc-pVTZ basis set (red bars) and the computational time relative to the cc-pVDZ basis set (blue squares). Calculations were performed using M06-2X at 298.15 K and 1 atm.

Papers I–III further confirm that the DFT binding energy is the largest source of errors in calculating Gibbs free energies for molecular clusters, and that the thermal contribution is well reproduced by density functional theory. In addition, we show in **Papers II** and **VI** that DFT with a small basis set, for instance 6-31++G**, is a cost-effective and accurate combination to obtain geometries and vibrational frequencies for atmospheric molecular clusters. However, a high-level correlation method is needed to correct inaccurate DFT electronic energies.

4.1.2 Applicability and Robustness of DLPNO–CCSD(T)

The accuracy of the calculated Gibbs free energies has been a concern in theoretical nucleation studies for several years. As discussed in the previous section, DFT even with a small basis set yields good geometries and thermal contributions, and the variation between different theory levels is relatively small.[368] In contrast, binding energies are very sensitive to the applied level of theory.[339] Different functionals often yield variation up to several kcal/mol, and since functionals are not systematically improvable, there is no a universal way to choose the best functional based on theory.

Canonical CCSD(T) calculations on atmospheric molecular clusters are valuable but limited to very small clusters due to the steep scaling of computational cost with respect to system

size. Obtaining information about larger clusters is essential to bridge the gap between theory and experiments. Therefore, a method which yields results close to the CCSD(T) method but which is applicable even for large clusters is needed. In **Papers I** and **III–VI** we have adopted a multi-step approach similar to Ortega *et al.*,[239] but instead of the RI-CC2 method we have utilized DLPNO–CCSD(T) for electronic energy corrections. In addition, instead of using B3LYP for optimization and frequencies, we have utilized three functionals (M06-2X, PW91, and ωB97X-D) which have shown to perform well for atmospheric molecular clusters.[332,333]

Paper I explores the applicability of the DLPNO–CCSD(T) method for atmospheric clustering studies. First we benchmarked DLPNO against the explicitly correlated canonical CCSD(T)-F12a method for small molecular clusters. Then we extended the analysis to medium-sized acid–base clusters. On the basis of these results, we further extended the study to large clusters and applied the DLPNO–CCSD(T) method to clusters consisting of up to five sulfuric acid and up to five ammonia or dimethylamine molecules. It should be emphasized that atmospheric acid–base clusters up to 10 molecules have previously been out of reach with accurate coupled cluster methods.

Table 3 shows the sulfuric acid dimer binding energies calculated using pure DFT, DLPNO–CCSD(T), and CCSD(T)-F12a. Binding energies vary from 15.9 kcal/mol to 21.1 kcal/mol when using DFT, but the scatter is significantly reduced when coupled cluster corrections are used. The DLPNO method systematically underbinds compared to the F12 method. Since the underbinding of DLPNO is consistent, DLPNO can be used as a lower bound for the CCSD(T) binding energies.

Table 3: Binding energies ΔE (kcal/mol) for the formation of sulfuric acid dimer calculated using DFT/6-311++G(3df,3pd), DLPNO–CCSD(T)/def2-QZVPP, and CCSD(T)-F12a/VDZ-F12 levels of theory. Table modified from **Paper I**.

	ΔE_{DFT}	$\Delta E_{\mathrm{DLPNO}}$	ΔE_{F12}
M06-2X	-19.1	-17.4	-17.9
PW91	-17.2	-17.0	-17.8
ωB97X-D	-18.4	-17.5	-17.9
B3LYP	-15.9	-17.2	-17.9
M06-2X-D	-19.3	-17.4	-17.9
CAM-B3LYP	-18.5	-17.3	-17.8
B3LYP-D	-19.7	-17.5	-18.0
CAM-B3LYP-D	-21.1	-17.4	-17.8

Paper I shows that DLPNO significantly reduces the scatter in the binding energies which are present in DFT calculations. Furthermore, the DLPNO error is a systematic underestimation compared to canonical coupled cluster. This error systematically depends on cluster size, which indicates that an empirical scaling factor can be derived. We calculated binding energies for small acid–base clusters and found the ratio between F12/DLPNO binding energies to be in the range of 1.01 to 1.04, with a mean ratio of 1.03. By scaling the DLPNO results by this factor the MAE is reduced from 1.3 to 0.3 kcal/mol, with a maximum error of 0.5 kcal/mol. It is worth noting that the F12/DLPNO scaling factors are not universal, as different systems behave differently in this respect. For example, in **Paper III**, we derived a scaling factor of 1.10 for ketodiperoxy acid–sulfuric acid clusters. It should be noted that the underbinding error might be due to basis set incompleteness error and too loose pair natural orbital criteria. DLPNO–CCSD(T) with an aug-cc-pVTZ basis set and a TightPNO criteria gives better results compared to the canonical CCSD(T) method. These energies are at least as good as the scaled binding energies, without applying empirical scaling.

Paper I extends the DLPNO calculations to large sulfuric acid–base clusters consisting of up to 10 molecules. In addition, we compared sulfuric acid dimer formation based on the DLPNO–CCSD(T)/def2-QZVPP//M06-2X/MG3S Gibbs free energies with previously published RI-CC2/aug-cc-pV(T+d)Z//B3LYP/CBSB7 results as well as experimental data obtained at the CLOUD chamber and flow tube experiments using the Atmospheric Cluster Dynamics Code (ACDC). The DLPNO//M06-2X level predicts lower formation of sulfuric acid dimers than both B3RICC2 and experiments performed at the CLOUD chamber for ammonia clusters. The agreement of results is much better for dimethylamine clusters, while B3RICC2 overestimates the dimer formation. In the case of flow tube experiments, DLPNO//M06-2X predicts significantly lower formation of sulfuric acid dimers for both ammonia and dimethylamine, but B3RICC2 yields good correspondence with the flow tube experiments. Although in worse agreement with the flow tube experiments, the DLPNO//M06-2X results should be more reliable than B3RICC2, as they come from a higher level of theory. The remaining errors are assumed to originate from other sources such as the rigid rotor-harmonic oscillator approximation, incomplete basis set, different global minima, and the effect of hydration.

Since basis set incompleteness might affect errors in correlated binding energy calculations, in **Paper VI** we have explored the basis set convergence of the DLPNO–CCSD(T) method. Test calculations were performed for small cluster formation reactions using cc-pVxZ and aug-cc-pVxZ basis sets, where x=2–5, and def2-xZVPP and ma-def2-xZVPP basis sets,

where x=3–4. We confirmed that the def2-QZVPP basis set offers good accuracy with low computational costs and it is a sufficient basis set for large molecular clusters (see Figure 15).

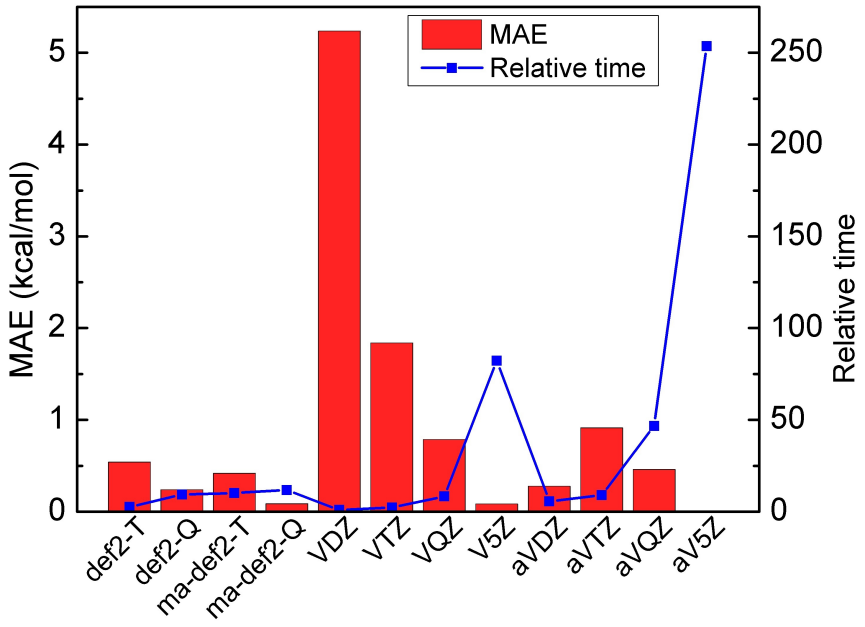

Figure 15: Mean absolute errors in the DLPNO–CCSD(T) binding energy (kcal/mol) relative to the aug-cc-pV5Z basis set (red bars) and the computational time relative to the cc-pVDZ basis set (blue squares).

Paper VI presents DLPNO–CCSD(T)/def2-QZVPP level of theory calculations for molecular clusters up to 105 atoms (6305 basis functions). Therefore, DLPNO calculations are run using local transformation (LT) type 3, which reduces memory requirements significantly without a loss in accuracy by calculating the local RI transformation on the fly. For instance, **Paper VI** shows that the required memory for the most expensive pair of resolution of identity transformation for the $(C_8H_{12}O_6)_2(H_2SO_4)$ cluster (59 atoms) is 102 GB when using LT1 and 22 GB when using LT2. LT3 reduces the memory cost even more, and for the largest 105 atoms cluster, 30 GB is a sufficient amount of memory per core. Overall, **Papers I**, **III**, and **VI** confirm that the DLPNO–CCSD(T) method is a reasonable choice for calculating binding energies for large atmospheric molecular clusters.

4.2 Applications

Highly oxidized organic compounds formed from terpenes are believed to play a major role in the formation and early growth of atmospheric aerosol particles.[125,126,128] **Papers III–VI** examine the strength of molecular interactions between oxidized organic compounds and sulfuric acid and evaluate the ability of oxidized organic compounds to act as a stabilizer in sulfuric acid induced clustering. Furthermore, they explore the effect of bases, ions, and water on the clustering of oxidized organic molecules and sulfuric acid. **Papers III–VI** apply a semiempirical technique for cluster structure sampling, M06-2X, PW91, and ωB97X-D functionals with 6-31++G** basis set (**Paper III** utilizes also the 6-311++G(3df,3pd) basis set) for geometries and frequencies, and DLPNO–CCSD(T)/def2-QZVPP level of theory for single point energies. Unless otherwise mentioned, Gibbs free energies are presented in kcal/mol and calculated using the RRHO approximation at 298.15 K temperature and 1 atm pressure.

4.2.1 Organic Peroxyacid Compounds

Terpenes can be oxidized rapidly in reactions initiated by an addition of OH radicals, NO_3, or O_3 to a double bond and a subsequent reaction with molecular oxygen.[130,131] Due to the complex molecular structure of terpenes, their autoxidation processes and specific structures of individual ELVOC species have not yet been fully resolved.[133] The most commonly used precursor for studying biogenic secondary organic aerosols is α-pinene $C_{10}H_{16}$, which has a structure consisting of a cyclohexene unit, a butyl ring, and three methyl groups (see Figure 16). In laboratory studies, cyclohexene C_6H_{10} is often used as a simpler model compound, as its oxidation chemistry partly resembles that of α-pinene. The formation of highly oxidized diperoxy acid compounds with additional keto and hydroperoxy substituents are reported from cyclohexene autoxidation.[129] As a proxy for monoterpene oxidation products, in **Papers III** and **IV**, we have evaluated the potential of ketodiperoxy acid $C_6H_8O_7$ to enhance sulfuric acid induced new-particle formation. Information about the clustering ability of this ketodiperoxy acid can further illuminate the direct involvement of monoterpene oxidation products in atmospheric clustering processes.

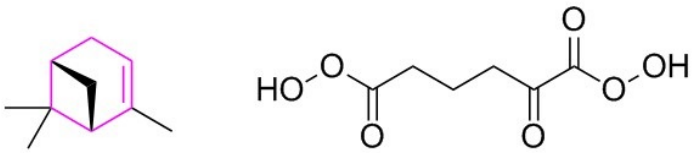

Figure 16: The structure of α-pinene, where its cyclohexene structure is marked in pink (left) and $C_6H_8O_7$ ketodiperoxy acid (right). Figure modified from **Paper III**.

Figure 17 shows the molecular structures of the $C_6H_8O_7$ dimer and the $(C_6H_8O_7)(H_2SO_4)$ complex. In **Paper III** we identified that $C_6H_8O_7$ interacts very weakly with itself and with sulfuric acid. This is due to the intramolecular hydrogen bonds in the ketodiperoxy acid monomer, which stabilize the isolated molecule with respect to its clusters. The formation of $(C_6H_8O_7)_2$ is thermodynamically unfavourable with Gibbs free energy of 3.1 kcal/mol. It involves breaking four intramolecular hydrogen bonds and forming four weaker intermolecular interactions. The low Gibbs free formation energy of $(C_6H_8O_7)(H_2SO_4)$ complex (ΔG=$-$0.2 kcal/mol) is also due to the thermodynamic stability of the ketodiperoxy acid monomer, which hinders the molecular interactions with sulfuric acid.

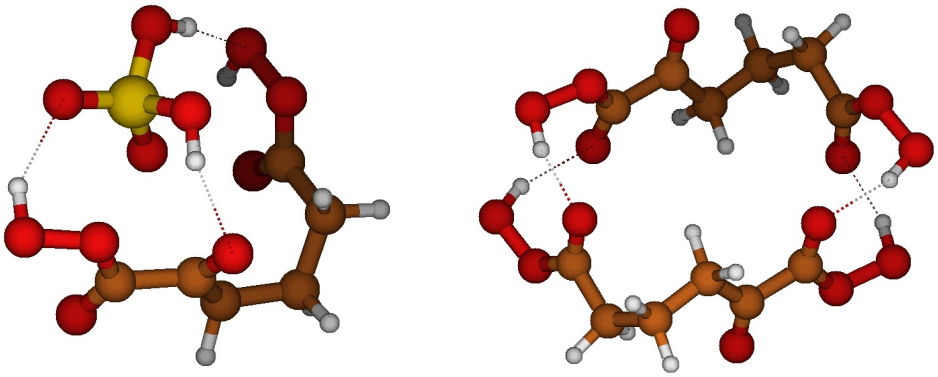

Figure 17: Ketodiperoxy acid–sulfuric acid complex $(C_6H_8O_7)(H_2SO_4)$ (left) and ketodiperoxy acid dimer $(C_6H_8O_7)_2$ (right). Colour coding: C=brown, O=red, S=yellow, and H=white.

We calculated the Gibbs free energies up to $(C_6H_8O_7)_2(H_2SO_4)_2$ cluster size and demonstrated that none of the formation pathways predict the formation of clusters containing two $C_6H_8O_7$. In addition, we studied the ability of ketodiperoxy acid to interact with aqueous sulfate ions. We showed that short-chained diperoxy acid compounds might contribute to aerosol growth, by partitioning into the aqueous aerosol phase. **Paper III** indicates that not only the oxygen-to-carbon ratio but also the number of strong hydrogen binding groups are important for

determining the ability of oxidized organic molecules to participate in atmospheric clustering, which is also confirmed by Kurtén *et al.*[132]

Paper IV continues the study of the molecular interaction between sulfuric acid and a ketodiperoxy acid compound by investigating the stabilizing effect of water, ammonia, and dimethylamine. Water and bases form hydrogen bonds to a vacant carbonyl group in the $C_6H_8O_7$ compound. The $(C_6H_8O_7)(H_2SO_4)(NH_3)$ and $(C_6H_8O_7)(H_2SO_4)((CH_3)_2NH)$ clusters exhibit a proton transfer from sulfuric acid to the bases which is not the case in the sulfuric acid–ammonia cluster. Figure 18 shows the $(C_6H_8O_7)(H_2SO_4)(X)$ cluster structures, where X=H_2O, NH_3, or $(CH_3)_2NH$. We found that the presence of water, ammonia, or dimethylamine enhances the molecular interaction between $C_6H_8O_7$ and H_2SO_4 compounds. However, the reaction free energies are only slightly negative, meaning that $C_6H_8O_7$ is likely to evaporate rapidly. The addition of a second ketodiperoxy acid to the $(C_6H_8O_7)(H_2SO_4)(X)$ cluster is thermodynamically unfavourable in all cases. Adding a second sulfuric acid to the $(C_6H_8O_7)(H_2SO_4)(X)$ cluster is favourable, but in all cases it is less favourable than the corresponding reaction without ketodiperoxy acid present. The addition of a second ketodiperoxy acid to the $(C_6H_8O_7)(H_2SO_4)_2(X)$ cluster is not favourable. Several other formation paths were considered, but none of them predict the formation of clusters containing two ketodiperoxy acids.

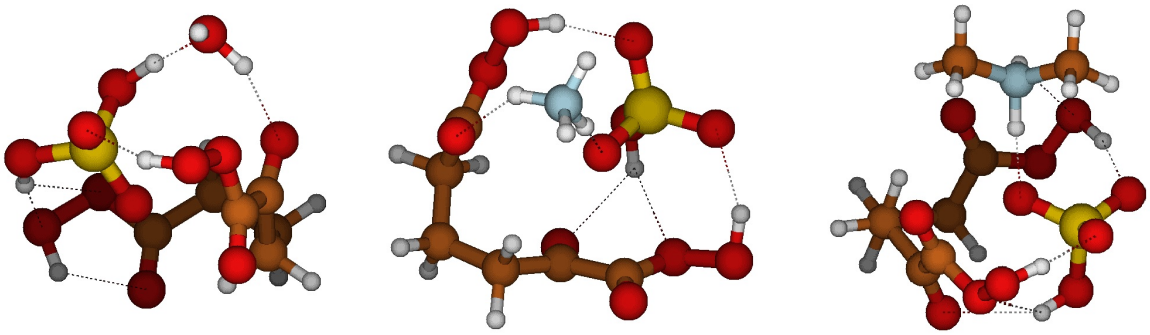

Figure 18: Molecular structures of $(C_6H_8O_7)(H_2SO_4)(H_2O)$ (left), $(C_6H_8O_7)(H_2SO_4)(NH_3)$ (middle), and $(C_6H_8O_7)(H_2SO_4)((CH_3)_2NH)$ (right) clusters. Colour coding: C=brown, O=red, S=yellow, N=blue, and H=white.

To further investigate the origin of the weak bonding of the peroxyacid moiety, we have compared the smaller performic and peracetic acids to formic and acetic acids. We analysed the natural bonding orbitals (NBOs) in order to get an indication of the hydrogen bond strength and the dimer formation energy. The occupation number of the antibonding OH orbital tells how much electron density is transferred to the opposing oxygen atom and into the hydrogen bond. When carboxylic acids form dimers, OH antibonding orbitals exhibit a significant increase in the occupation numbers. This favours dimer formation by strengthening intermolecular hydrogen bonds. In the case of peroxyacids, the occupation number of the OH antibonding orbital is higher for monomer than for dimer. This is because of the intramolecular hydrogen bonds in monomers are stronger than the intermolecular hydrogen bonds in the dimer structure. In addition, we investigated the ability of formic and performic acids to donate and accept protons in order to get an indicator of the strength of intermolecular hydrogen bonds. We used a chloride ion for probing the hydrogen bond donor strength and hydrogen chloride to probe the hydrogen bond acceptor strength. We demonstrated that the strength of hydrogen bond donor is identical for formic and performic acids. However, the hydrogen bond acceptor strength is four times lower for performic acid.

Paper IV confirms that strong intramolecular hydrogen bonds in the peroxyacid monomer reduce the potential of peroxyacids to participate in dimer or cluster formation and that the weak clustering ability of the peroxyacid compounds is due to the lack of strong hydrogen bond acceptors. The participation of water, ammonia, or dimethylamine does not promote the interaction enough to make ketodiperoxy acid–sulfuric acid clustering occur under atmospheric conditions. Thus we concluded that autoxidation products consisting mainly of peroxyacid, hydroperoxide, and carbonyl groups cannot have an important role in the first steps of new-particle formation. Thereby they may only contribute to aerosol mass in the subsequent growth of freshly nucleated particles. We showed that the oxygen-to-carbon ratio alone cannot be used as a proxy for the volatility of oxidized organic compounds and hence their potential to enhance cluster formation. **Paper IV** indicates that oxidation products with carboxylic acid groups would be better candidates to participate in atmospheric clustering, since they are able to form stronger intermolecular hydrogen bonds. This is further confirmed in **Paper X**.

4.2.2 Multi-Carboxylic Acids

Paper V examines the molecular interactions between 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA, identified from α-pinene oxidation) and sulfuric acid up to cluster sizes of $(MBTCA)_3(H_2SO_4)_3$. The formation of the $(MBTCA)(H_2SO_4)$ complex occurs by forming three hydrogen bonds, for which the reaction free energy is -6.2 kcal/mol. The interaction between MBTCA and sulfuric acid is slightly more favourable than the interaction between sulfuric acid molecules (ΔG=-5.2 kcal/mol) or sulfuric acid and ammonia (ΔG=-4.6 kcal/mol). The formation of $(MBTCA)(H_2SO_4)$ is still less favourable than the formation of the sulfuric acid–dimethylamine cluster, for which the reaction free energy is -10.6 kcal/mol. The interaction between two MBTCA molecules is weak (ΔG=-2.2 kcal/mol), and thus the formation of the heterodimer between sulfuric acid and MBTCA is more favourable than the formation of either homodimers. Figure 19 shows $(MBTCA)(H_2SO_4)$ and $(MBTCA)_2$ cluster structures.

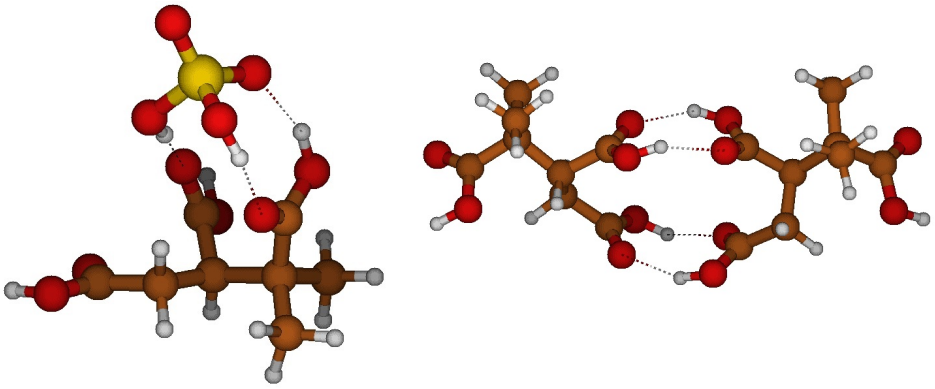

Figure 19: Molecular structures of $(MBTCA)(H_2SO_4)$ (left) and $(MBTCA)_2$ (right) clusters. Colour coding: C=brown, O=red, S=yellow, and H=white.

The first two additions of MBTCA compounds to the $(MBTCA)(H_2SO_4)$ heterodimer are thermodynamically more favourable than the corresponding first two additions of sulfuric acid molecules. This is because a higher cluster stabilization is achieved when the amount of sulfuric acid–carboxylic acid interactions is maximized. The $(MBTCA)_2(H_2SO_4)$ cluster is able to form seven hydrogen bonds, while $(MBTCA)(H_2SO_4)_2$ forms six. Also $(MBTCA)_3(H_2SO_4)$ can form nine hydrogen bonds, but $(MBTCA)(H_2SO_4)_3$ only forms eight. The addition of a second sulfuric acid to the $(MBTCA)_2(H_2SO_4)$ cluster is particularly favourable (ΔG=-9.9 kcal/mol), and the addition of MBTCA to the $(MBTCA)(H_2SO_4)_2$ cluster is even more

favourable (ΔG=−11.6 kcal/mol). The stability of the $(MBTCA)_2(H_2SO_4)_2$ cluster is comparable to the sulfuric acid–dimethylamine complex, and thus it might be relatively stable against re-evaporation. The $(MBTCA)_2(H_2SO_4)_2$ cluster is able to form ten hydrogen bonds, with eight of them connecting between sulfuric acid–carboxylic acid groups, and the remaining two between carboxylic acid groups.

The $(MBTCA)_3(H_2SO_4)_2$ cluster forms 12 hydrogen bonds, and it can be formed by adding a sulfuric acid molecule to the $(MBTCA)_3(H_2SO_4)$ cluster (ΔG=−10.7 kcal/mol), or by adding an MBTCA molecule to the $(MBTCA)_2(H_2SO_4)_2$ cluster (ΔG=−5.6 kcal/mol). Also the $(MBTCA)_2(H_2SO_4)_3$ cluster has a total of 12 hydrogen bonded interactions, but in contrast to the $(MBTCA)_3(H_2SO_4)_2$ cluster, they all are between sulfuric acid and MBTCA. Therefore, both pathways for forming the $(MBTCA)_2(H_2SO_4)_3$ cluster by addition of sulfuric acid or MBTCA monomer to the pre-existing cluster are highly favourable, with reaction free energies of −10.6 kcal/mol and −18.4 kcal/mol, respectively. The formation of $(MBTCA)_2(H_2SO_4)_4$ or $(MBTCA)_3(H_2SO_4)_3$ clusters by adding sulfuric acid to $(MBTCA)_2(H_2SO_4)_3$ or $(MBTCA)_3(H_2SO_4)_2$ clusters yield high reaction free energies, because the reactant clusters are very stable. The higher formation free energies for forming larger clusters with more than three sulfuric acid and three MBTCA molecules could mean that more sulfuric acid molecules are needed to maximize the amount of sulfuric acid–carboxylic acid interactions. As three MBTCA monomers have a total of nine carboxylic acid groups, it indicates that 4–5 sulfuric acid molecules are required to maximize the amount of direct hydrogen bonded interactions. Figure 20 summarizes the reaction free energies of forming sulfuric acid–MBTCA clusters.

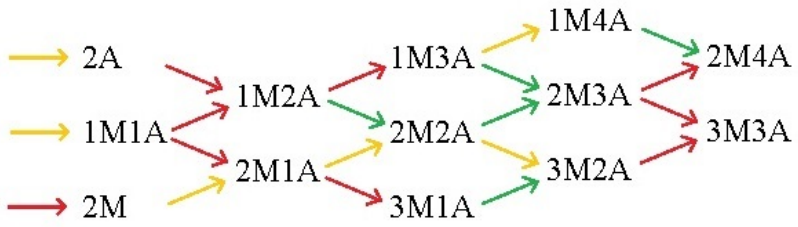

Figure 20: Gibbs free energy diagram for MBTCA–sulfuric acid clusters at 298.15 K and 1 atm. M represents MBTCA and A is sulfuric acid. Colour coding: red $>$ −5 kcal/mol, yellow −5 to −10 kcal/mol, and green $<$ −10 kcal/ mol. Figure adapted from **Paper VI**.

In **Paper V**, we have obtained the actual Gibbs free energy surface of the sulfuric acid–MBTCA clusters at atmospheric concentrations of $[H_2SO_4]=10^{7}$ cm^{-3} and $[MBTCA]=10^{7}$

cm^{-3} at 298.15 K from the law of mass action. For any given cluster there is no growth direction that leads to a lower Gibbs free formation energy, *i.e.*, the addition of either MBTCA or sulfuric acid molecules always leads to a higher formation free energy. Figure 21 shows that the Gibbs free energy increases towards the system boundaries meaning that the growth within the system is unfavourable. There does not exist a critical cluster within the simulation box.

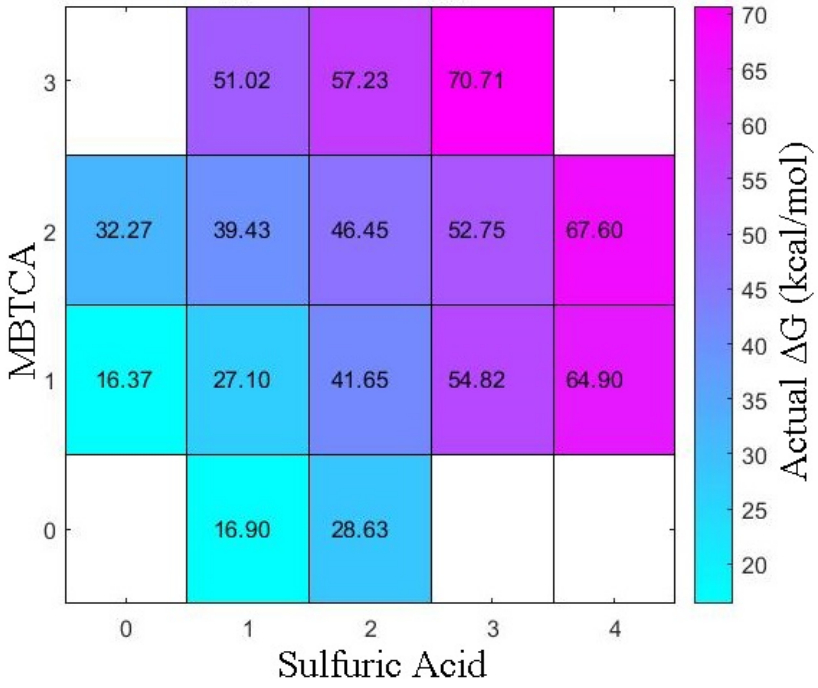

Figure 21: Actual free energy surface (kcal/mol) of the MBTCA–sulfuric acid grid at 298.15 K, when [sulfuric acid]=10^{7} cm^{-3} and [MBTCA]=10^{7} cm^{-3}. Figure modified from **Paper V**.

To form new particles the collision rate of monomers to the clusters must exceed the cluster evaporation rates beyond some cluster size. We calculated the ratio between the sulfuric acid or MBTCA monomer collision rates and the total evaporation rate. For all cases the ratio is below 1, which indicates that clusters evaporate faster than they collide with sulfuric acid or MBTCA monomers. The $(MBTCA)_2(H_2SO_4)_2$ and $(MBTCA)_2(H_2SO_4)_3$ clusters are more stable against evaporation than all other clusters. While these clusters cannot grow by further addition of sulfuric acid or MBTCA, they could act as seeds for further growth by taking up other stabilizing vapour molecules. **Paper V** leads to the conclusion that MBTCA and sulfuric acid by themselves could not drive new-particle formation under realistic conditions in the atmospheric lower boundary layer.

Paper VI examines the effect of bisulfate HSO_4^-, ammonium NH_4^+, and ammonia NH_3 on the clustering of sulfuric acid and pinic acid or MBTCA. Ions may contribute to the clustering process by stabilizing condensing species from evaporating. Bisulfate and ammonium are believed to be key participants in ion-induced nucleation. Ammonia, for instance, is able to stabilize larger sulfuric acid or organic acid containing particles. We have investigated the ability of ammonia and ions to enhance cluster formation and growth by decreasing the overall evaporation rates of the clusters containing sulfuric acid and multi-carboxylic acids.

We have studied pinic acid clusters up to the $(C_9H_{14}O_4)_2(H_2SO_4)_2(X)_1$ size, where X=HSO_4^-, NH_3, or NH_4^+. When bisulfate or ammonium is present, the Gibbs free binding energies are about 20 kcal/mol more negative compared to the two-component sulfuric acid–pinic acid clusters. In the case of two-component pinic acid–sulfuric acid clusters, none of the reaction steps is very favourable. The interaction between bisulfate and sulfuric acid is very strong, and thus the addition of pinic acid to the sulfuric acid–bisulfate clusters is unfavourable. Bisulfate-containing clusters can more likely grow via the pinic acid–bisulfate cluster, which is stabilized by two hydrogen bonds with bisulfate and carboxylic acid groups. Ammonia interacts weakly with pinic acid or sulfuric acid, and none of the formation routes are thermodynamically favourable. Only one hydrogen bond is formed between ammonia and pinic acid. The interaction between ammonium and pinic acid or sulfuric acid is strong due to the formation of two hydrogen bonds. Pinic acid is able to form strong, nearly linear hydrogen bonds with ammonium, whereas sulfuric acid forms hydrogen bonds with angles of 140°.

Figure 22 shows the overall evaporation rates, *i.e.*, the sum over all evaporation and fragmentation processes ($\sum \gamma$) at 273.15 K. All evaporation rates are high, with the exception of the sulfuric acid–bisulfate clusters and the pinic acid–ammonium cluster. Therefore, the growth of pinic acid containing clusters is very unlikely, which has also been shown in the previous study of the neutral sulfuric acid–pinic acid clusters by Elm *et al.*[391] The evaporation rates of two-component sulfuric acid–pinic acid clusters are usually lower than those of corresponding bisulfate, ammonia, or ammonium containing clusters. All clusters containing both sulfuric acid and bisulfate are evaporating towards two-component sulfuric acid–bisulfate clusters since their interaction is significantly stronger than any other interaction. The ammonium ion interacts strongly with carboxylic acid groups, and therefore, ammonium-containing three-component clusters evaporate towards two-component pinic acid–ammonium clusters.

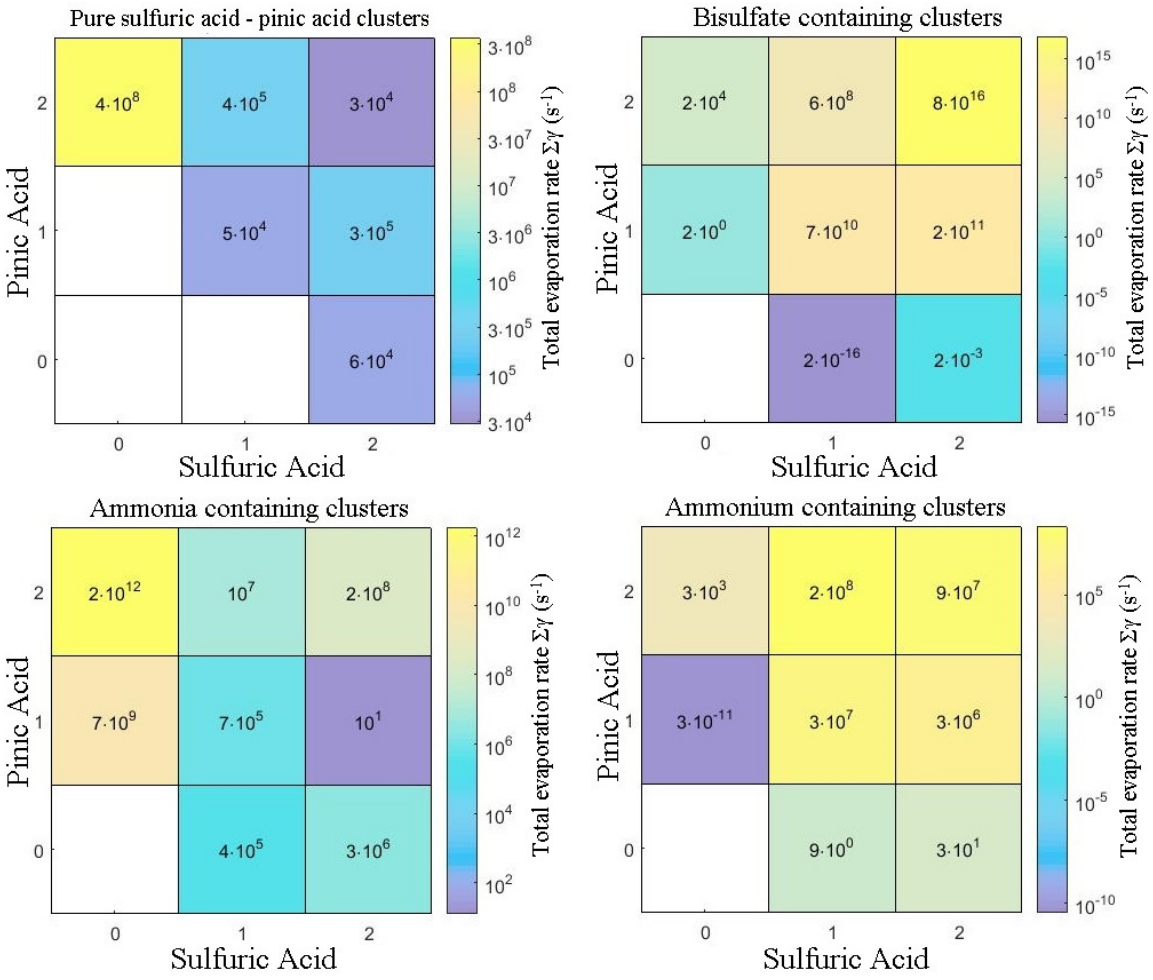

Figure 22: Overall evaporation rates $\sum \gamma$ (sum over all decomposition processes) for pinic acid clusters at 273.15 K. Figure adapted from **Paper VI**.

Figure 23 presents the calculated Gibbs free energies for MBTCA containing clusters up to the size of $(C_8H_{12}O_6)_3(H_2SO_4)_3(X)_1$, where X=$HSO_4^-$, NH_3, or NH_4^+. Bisulfate and ammonium ions bind strongly with MBTCA by forming hydrogen bonds, but the interaction between ammonia and MBTCA is weak. Bisulfate and ammonium are reducing the Gibbs free formation energies by 20–40 kcal/mol compared to the two-component MBTCA–sulfuric acid clusters. Since the interaction with sulfuric acid and bisulfate or ammonium is strong, the low Gibbs free energy values are mainly originating from the interaction between sulfuric acid and ions, especially in the case of bisulfate.

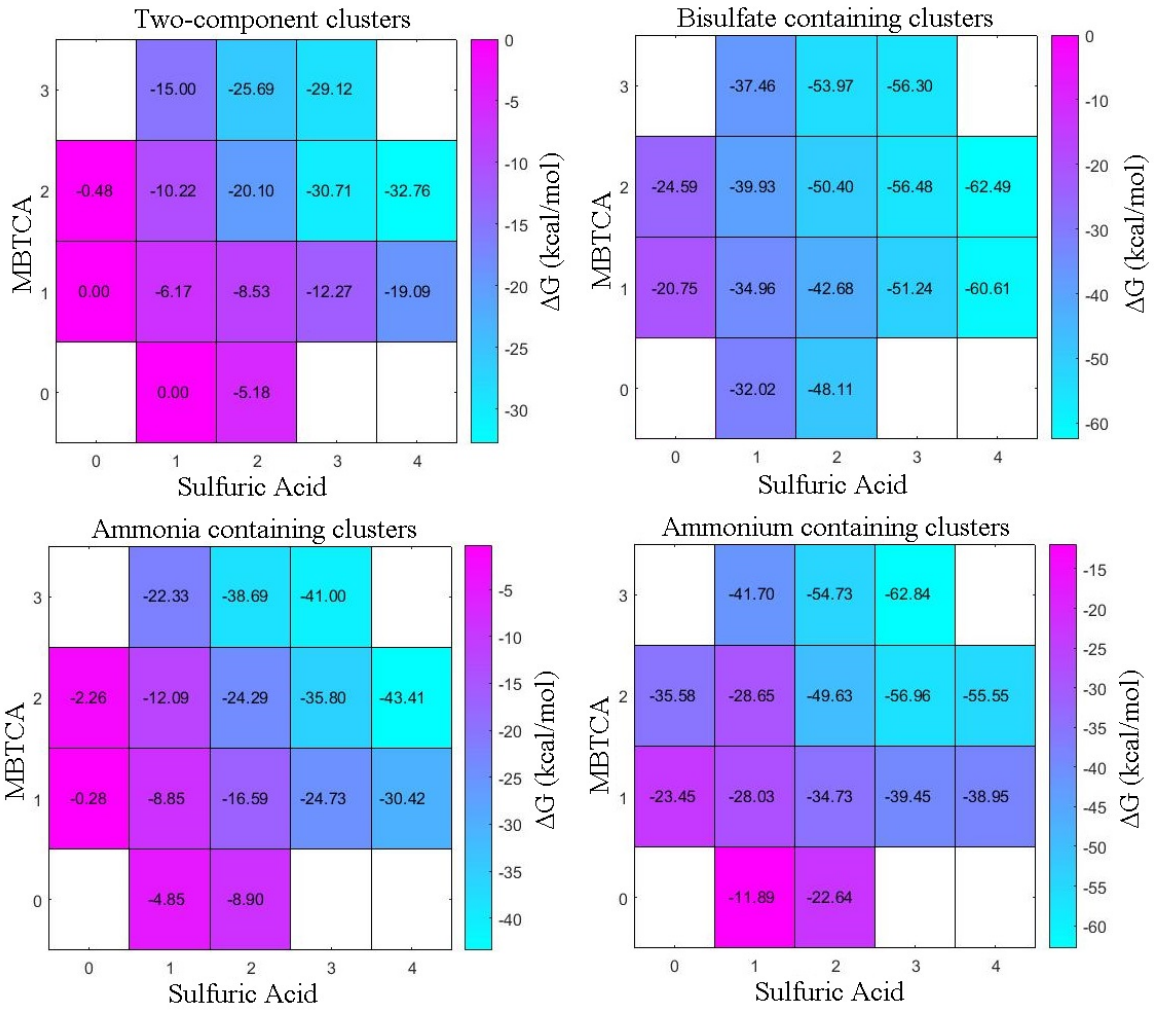

Figure 23: Gibbs free formation energies (kcal/mol) for MBTCA clusters at 298.15 K.

The overall evaporation rates for MBTCA clusters at 273.15 K are presented in Figure 24. Similarly to the case of pinic acid, bisulfate increases the evaporation rates of the three-component clusters due to the very high stability of two-component sulfuric acid–bisulfate clusters. Therefore, all three-component clusters evaporate fast towards sulfuric acid–bisulfate clusters. The presence of ammonia or ammonium can either increase or decrease the evaporation rates by several orders of magnitude. The two-component MBTCA–ammonium clusters are particularly stable against evaporation because the interaction between ammonium and carboxylic acid groups is strong. The most stable three-component cluster consists of one ammonia, three MBTCA, and two sulfuric acid molecules, and if it is able to form, it might act as a seed for addition of other stabilizing vapour molecules.

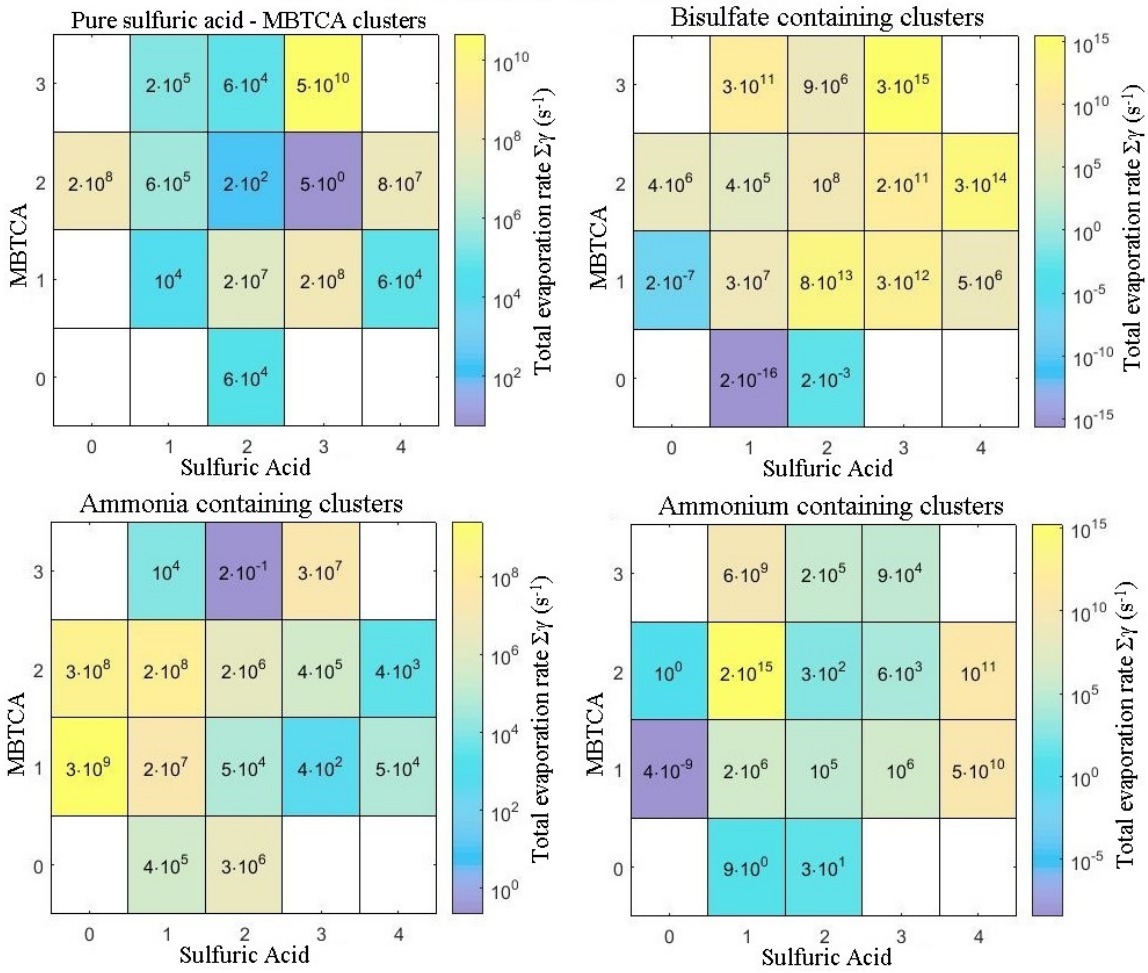

Figure 24: Overall evaporation rates $\sum \gamma$ (sum over all decomposition processes) for MBTCA clusters at 273.15 K. Figure adapted from **Paper VI**.

Paper VI shows that under atmospheric conditions and realistic vapour pressures it is unlikely that organic acid containing clusters can grow into large stable clusters. Thus we concluded that organic acids and sulfuric acid even together with bisulfate, ammonia, or ammonium cannot drive the observed new-particle formation events via clustering mechanisms. However, experimental studies have found oxidized organic compounds to participate in the initial steps of atmospheric new-particle formation, especially via ion-induced pathways.[97,126,128] In addition, we showed in **Paper X** that multi-carboxylic acids are the most prominent candidates of oxidized organic compounds to form stable clusters with sulfuric acid. Therefore, the results of **Paper VI** lead to the conclusion that some other mechanisms are required to explain experimentally observed formation events (see Section 5 for possible explanations).

4.3 Overview of Papers and the Author's Contribution

Paper I

We utilized a domain-based local pair natural orbital coupled cluster with singles, doubles, and perturbative triples (DLPNO–CCSD(T)) method for atmospheric acid–base clustering. We calculated binding energies for a set of small to medium-sized acid–base clusters using DFT, DLPNO–CCSD(T)/def2-QZVPP, and CCSD(T)-F12/VDZ-F12 levels of theory. We confirmed that the DLPNO–CCSD(T) results are much more reliable than DFT results, but they yield a systematic underbinding for studied systems. Therefore, we applied a scaling factor of 1.03 to the DLPNO–CCSD(T) binding energies in order to reduce the mean absolute error from 1.3 kcal/mol down to 0.3 kcal/mol compared to the reference CCSD(T)-F12 values. This novel approach solved the previous problem related to inaccurate energies, originating from pure DFT approach or energy corrections calculated using the RI-CC2 method. In fact, this was the first study where a method close to the canonical CCSD(T) accuracy was employed to large acid–base clusters with up to 10 molecules. The author performed all coupled cluster benchmark calculations for small complexes and compared DFT and CC methods in the case of medium-size clusters. The author extended the study for large acid–base clusters and wrote the first draft of that part of the paper.

Paper II

We studied the basis set convergence of density functionals with respect to the binding energy, the thermal contribution to the Gibbs free energy, and the optimized geometry. Our test set contained six hydrogen-bonded cluster formation reactions which represent some of the most important non-covalent interactions in molecular clusters. To further investigate whether the findings for the small complexes are also valid for larger clusters, we extended the calculations for clusters up to four acid and four base molecules. We utilized three density functionals which are commonly used in atmospheric clustering studies with correlation consistent, Pople-type, and polarization consistent basis sets with different amounts of diffuse and polarization functions. We showed that partially augmented basis sets yield as accurate results as fully augmented basis sets at lower computational expense, and that the small 6-31++G(d,p) basis set is sufficient for obtaining geometries and frequencies of atmospheric molecular clusters. In addition, we studied the effects of anharmonic and quasi-harmonic corrections on the thermal contribution to the Gibbs free energy. We found the vibrational anharmonic corrections to be relatively small, but the quasi-harmonic corrections yield for large clusters several kcal/mol difference compared to the RRHO approximation. The author

performed all quantum chemical calculations except anharmonic ones, introduced and applied the quasi-harmonic approximation as the first for molecular clusters, and wrote the paper.

Paper III

We studied the molecular interactions between sulfuric acid and a ketodiperoxy acid $C_6H_8O_7$ formed via autoxidation of cyclohexene. We showed that $C_6H_8O_7$ interacts very weakly with both itself and sulfuric acid because of strong intramolecular hydrogen bonds in the peroxyacid groups of $C_6H_8O_7$. We concluded that the oxygen-to-carbon ratio cannot solely be used as a proxy for volatility in clusters involving oxidized organic compounds since the number of strong hydrogen binding groups is equally important. This study was the very first where DLPNO–CCSD(T) method was utilized to evaluate the stability of atmospheric clusters. The author performed all coupled cluster benchmark calculations, estimated approximate DLPNO–CCSD(T) binding free energies for the interaction between aqueous sulfate and ($C_6H_8O_7$) in a water solvent, and applied the DLPNO–CCSD(T) method for ketodiperoxy acid–sulfuric acid clusters.

Paper IV

We evaluated the ability of water, ammonia, and dimethylamine to stabilize sulfuric acid and $C_6H_8O_7$ ketodiperoxy acid clusters. We found that the presence of water or base molecules enhances the molecular interaction between a single ketodiperoxy acid and sulfuric acid. The addition of a second ketodiperoxy acid to the cluster is thermodynamically unfavourable in all cases. To further investigate the origin of the weak binding of peroxyacid compounds we utilized atoms in molecules (AIM) and natural bonding orbital (NBO) analysis. The weak molecular interaction is caused by the lack of a strong hydrogen bond acceptor and the formation of a strong intramolecular hydrogen bond in peroxyacid monomer structure. We concluded that autoxidation products containing only peroxyacid, hydroperoxide, and carbonyl groups cannot be key species in the first steps of new-particle formation. The author was responsible for all of the DLPNO–CCSD(T) electronic energy corrections for ketodiperoxy acid and sulfuric acid clusters with water, ammonia, and dimethylamine.

Paper V

We investigated the formation of atmospheric clusters consisting of sulfuric acid and the α-pinene oxidation product 3-methyl-1,2,3-butanetricarboxylic acid (MBTCA). We found that the formation of the sulfuric acid–MBTCA heterodimer is more favourable than either of the sulfuric acid or MBTCA homodimers. The molecular interaction between MBTCA and sulfu-

ric acid is thermodynamically favourable and a large stabilization of the cluster is achieved when the amount of sulfuric acid–carboxylic acid interactions is maximized. We showed that clusters consisting of 2–3 MBTCA and 2–3 sulfuric acid molecules are particularly stable and their stability is comparable to the sulfuric acid–dimethylamine cluster. In order to evaluate the stability of clusters under atmospheric conditions and realistic vapour pressures, we obtained the actual Gibbs free energy surface of the clusters from the law of mass action and calculated the ratio of the rate of collisions with sulfuric acid or MBTCA molecules to the total evaporation rate of each cluster. Cluster kinetics calculations showed that the cluster growth is limited by a weak formation of the largest sulfuric acid–MBTCA clusters studied. The author performed all DLPNO–CCSD(T) calculations for MBTCA–sulfuric acid clusters.

Paper VI
We evaluated how bisulfate, ammonia, and ammonium affect the clustering of organic acids and sulfuric acid. We showed that the presence of ions enhances the first steps of cluster formation as bisulfate stimulates the clustering through the addition of sulfuric acid and ammonium instead of the addition of organic acids. At atmospheric conditions, however, further cluster growth is limited due to the weak interaction and fast evaporation of the larger three-component clusters. Therefore, it is unlikely that organic multi-carboxylic acids and sulfuric acid, even together with bisulfate, ammonia, or ammonium can drive new-particle formation via clustering mechanisms. We suggested that other mechanisms such as chemical reactions are required to explain observed new-particle formation events in the presence of oxidized organic compounds. The author carried out the cluster sampling, benchmarking, quantum chemical calculations and cluster kinetics. The author wrote the paper.

5 Future Perspectives

An overview of the key findings and conclusions of **Papers I–VI** is given in previous section. The research presented in this thesis demonstrates that quantum chemistry is a powerful tool for studying the molecular-level formation mechanisms and stabilities of atmospheric clusters. This thesis addresses that care must be taken when choosing the computational method for a given problem, as calculations performed with inadequate methods often lead to erroneous conclusions. The robustness of the DLPNO–CCSD(T)//DFT level is studied and the accuracy and the applicability for large molecular clusters is confirmed. State-of-the-art quantum chemical methodologies are presented, and recommended to be applied for future

clustering studies. In addition, the quasi-harmonic approach to correct the failure of the RRHO approximation in the case of weakly-bound molecular clusters is presented. We suggested extending the study of the accuracy of the quasi-harmonic approximation to a wider test set, because the low-lying frequencies might be a significant error source (up to several kcal/mol) for large molecular clusters.

The applications presented in this thesis offer a computational aspect of sulfuric acid clustering with oxidized organic compounds. The results indicate that non-basic organic compounds are unlikely to have a strongly enhancing role in the initial steps of sulfuric acid driven new-particle formation in atmospheric conditions via a clustering mechanism. However, experimental studies have shown that oxidized organic compounds participate in the initial steps of new-particle formation.[126,128] Due to the disagreement between theoretical and experimental findings, some other mechanisms or compounds are needed to explain the experiments.

Chemical Reactions

In addition to non-covalent interactions, also chemical reactions should be considered when studying atmospheric new-particle formation involving oxidized organic compounds. One possible reason for the discrepancy between experimental and theoretical results might be the formation of covalently-bound dimers or organosulfates.[392–394] Both field and smog chamber measurements have found that SOAs contain oligomeric macromolecules, which might be formed in either gas- or particle-phase reactions.[395–398] A large number of different monomers are present in the atmosphere, which means that even more dimer products exist. Reaction kinetics and thermodynamics of dimer or organosulfate formation as well as the relative significance of participating compounds are unclear. Therefore, computational studies are needed to augment measurements in order to probe the detailed reaction mechanisms and molecular structures. The formed dimer or organosulfate products very likely have a lower saturation vapour pressure than the reacting monomers due to a higher molecular mass and a larger number of functional groups.[132,399] These clusters would be more stable against evaporation; therefore, cluster-phase reactions might play a significant role in atmospheric new-particle formation and growth.

Strong Bases

It has been demonstrated that strong bases with low abundance might be key players in the first steps of sulfuric acid driven particle formation.[92] Future studies on, for instance, guanidine ($HNC(NH_2)_2$) and its derivatives might be interesting research subjects in atmospheric sciences. Guanidine is a strong organobase, which has been found in urine as a normal product

of protein metabolism.[400,401] It can also be emitted from anthropogenic sources such as plastic and explosive production.[402,403] The cation form of guadine is guanidinium, which is extraordinarily stable due to the Y-aromaticity, plane symmetry, and resonance stabilization.[404] The molecular interaction of the sulfuric acid–guanidine complex is four times stronger than sulfuric acid–ammonia and almost two times stronger than sulfuric acid–dimethylamine complexes (see Figure 25). Therefore, sulfuric acid–guanidine clusters could act as seeds for further growth via the uptake of other vapour molecules, for instance, more abundant amines with lower basicity or oxidized organic compounds. This implies that a wide range of different chemical species is required to explain the atmospheric particle formation events. On that account, it is important to uncover the central compounds driving the cluster formation in different environments.

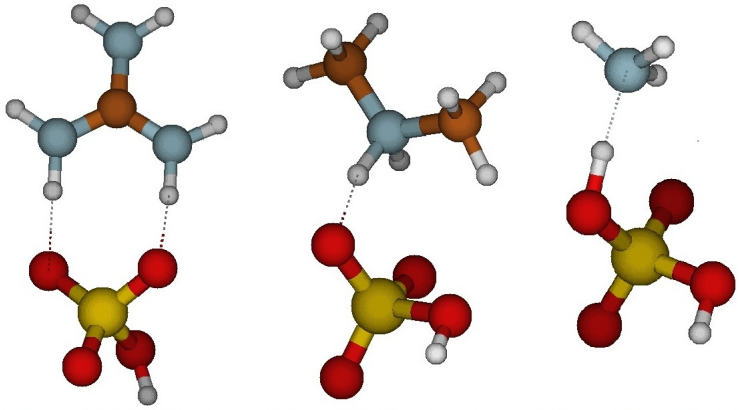

Figure 25: Molecular structures and Gibbs free formation energies (DLPNO–CCSD(T)/aug-cc-pVTZ//M06-2X/6-31++G**) at 298.15 K and 1 atm for sulfuric acid complexes with guanidine (left), dimethylamine (middle), and ammonia (right). Colour coding: C=brown, O=red, S=yellow, N=blue, and H=white.

Concluding Remarks

Further studies — both experimental and theoretical — are needed to explore the importance of oxidized organic compounds in the new-particle formation in the atmosphere. In order to elucidate the exact cluster formation mechanisms, the role of chemical reactions and the stabilizing effect of strong bases must be taken into consideration. The quantum chemical approaches presented in this thesis offer a cost-effective way to produce state-of-the-art thermodynamic data for clustering studies. Highly accurate electronic structure calculations are required to get a better understanding of molecular-level cluster formation mechanisms in the atmosphere.

References

[1] W. Hinds, *Aerosol Technology: Properties, Behavior, and Measurement of Airborne Particles*, John Wiley and Sons, Inc., New York, NY, USA, 2nd edn., 1999.

[2] A. Virtanen, J. Joutsensaari, T. Koop, J. Kannosto, P. Yli-Pirilä, J. Leskinen, J. M. Mäkelä, J. K. Holopainen, U. Pöschl, M. Kulmala, D. R. Worsnop and A. Laaksonen, *Nature*, 2010, **467**, 824–827.

[3] D. V. Spracklen, K. S. Carslaw, M. Kulmala, V.-M. Kerminen, G. W. Mann and S.-L. Sihto, *Atmos. Chem. Phys.*, 2006, **6**, 5631–5648.

[4] P. Mönkkönen, I. Koponen, K. Lehtinen, K. Hämeri, R. Uma and M. Kulmala, *Atmos. Chem. Phys.*, 2005, **5**, 57–66.

[5] J. H. Seinfeld and S. N. Pandis, *Atmospheric Chemistry and Physics: From Air Pollution to Climate Change*, John Wiley & Sons, 2016.

[6] P. Kulkarni, P. A. Baron and K. Willeke, *Aerosol Measurement: Principles, Techniques, and Applications*, John Wiley & Sons, 2011.

[7] J. Aitken, *Nature*, 1888, **37**, 428–430.

[8] J. Aitken, *Nature*, 1880, **23**, 195–197.

[9] J. Merikanto, D. V. Spracklen, G. W. Mann, S. J. Pickering and K. S. Carslaw, *Atmos. Chem. Phys.*, 2009, **9**, 8601–8616.

[10] M. Dal Maso, M. Kulmala, I. Riipinen, R. Wagner, T. Hussein, P. P. Aalto and K. E. Lehtinen, *Boreal Env. Res.*, 2005, **10**, 323.

[11] A. Nel, *Science*, 2005, **308**, 804–806.

[12] V. Ramanathan, P. Crutzen, J. Kiehl and D. Rosenfeld, *Science*, 2001, **294**, 2119–2124.

[13] H.-E. Wichmann and A. Peters, *Phil. Trans. R. Soc. A Math. Phys. Eng. Sci.*, 2000, **358**, 2751–2769.

[14] D. Stieb, S. Judek and R. Burnett, *J. Air Waste Manage. Assoc.*, 2002, **52**, 470–484.

[15] B. Brunekreef and S. T. Holgate, *Lancet*, 2002, **360**, 1233–1242.

[16] A. J. Cohen, H. Ross Anderson, B. Ostro, K. D. Pandey, M. Krzyzanowski, N. Künzli, K. Gutschmidt, A. Pope, I. Romieu, J. M. Samet *et al.*, *J. Toxicol. Environ. Health, Part A*, 2005, **68**, 1301–1307.

[17] X. Tie, D. Wu and G. Brasseur, *Atmos. Environ.*, 2009, **43**, 2375–2377.

[18] M. Kampa and E. Castanas, *Environ. Pollut.*, 2008, **151**, 362–367.

[19] K. R. Smith, N. Bruce, K. Balakrishnan, H. Adair-Rohani, J. Balmes, Z. Chafe, M. Dherani, H. D. Hosgood, S. Mehta, D. Pope *et al.*, *Annu. Rev. Public Health*, 2014, **35**, 185–206.

[20] E. Saikawa, V. Naik, L. W. Horowitz, J. Liu and D. L. Mauzerall, *Atmos. Environ.*, 2009, **43**, 2814–2822.

[21] WHO (World Health Organization), *Public health, environmental and social determinants of health*, 2014.

[22] N. Parks, *Environ. Sci. Technol.*, 2009, **43**, 8475–8476.

[23] J. Lelieveld, *Nature*, 2006, **443**, 405–406.

[24] Y. Malhi, J. T. Roberts, R. A. Betts, T. J. Killeen, W. Li and C. A. Nobre, *Science*, 2008, **319**, 169–172.

[25] M. C. Serreze, M. M. Holland and J. Stroeve, *Science*, 2007, **315**, 1533–1536.

[26] B. Hansen, S. Østerhus, D. Quadfasel and W. Turrell, *Science*, 2004, **305**, 953–954.

[27] J. Hansen, M. Sato, R. Ruedy, P. Kharecha, A. Lacis, R. Miller, L. Nazarenko, K. Lo, G. Schmidt, G. Russell *et al.*, *Atmos. Chem. Phys.*, 2007, **7**, 2287–2312.

[28] V. Ramanathan and Y. Feng, *Atmos. Environ.*, 2009, **43**, 37–50.

[29] S. A. Montzka, E. J. Dlugokencky and J. H. Butler, *Nature*, 2011, **476**, 43.

[30] IPCC (Intergovernmental Panel on Climate Change), *Climate change 2013: The Physical Science Basis*, 2013.

[31] G. Myhre, *Science*, 2009, **325**, 187–190.

[32] H. Yu, Y. Kaufman, M. Chin, G. Feingold, L. Remer, T. Anderson, Y. Balkanski, N. Bellouin, O. Boucher, S. Christopher *et al.*, *Atmos. Chem. Phys.*, 2006, **6**, 613–666.

[33] S. Twomey, *Atmos. Environ.*, 1991, **25A**, 2435–2442.

[34] J. Pierce and P. Adams, *Atmos. Chem. Phys. Discuss.*, 2008, **8**, 16291–16333.

[35] D. Rosenfeld, *Science*, 2000, **287**, 1793–1796.

[36] D. V. Spracklen, K. S. Carslaw, M. Kulmala, V.-M. Kerminen, S.-L. Sihto, I. Riipinen, J. Merikanto, G. W. Mann, M. P. Chipperfield, A. Wiedensohler, W. Birmili and H. Lihavainen, *Geophys. Res. Lett.*, 2008, **35**, L06808.

[37] D. Rosenfeld, U. Lohmann, G. B. Raga, C. D. O'Dowd, M. Kulmala, S. Fuzzi, A. Reissell and M. O. Andreae, *Science*, 2008, **321**, 1309–1313.

[38] J. E. Penner, J. Quaas, T. Storelvmo, T. Takemura, O. Boucher, H. Guo, A. Kirkevåg, J. E. Kristjánsson and Ø. Seland, *Atmos. Chem. Phys.*, 2006, **6**, 3391–3405.

[39] U. Lohmann and J. Feichter, *Atmos. Chem. Phys.*, 2005, **5**, 715–737.

[40] J. Haywood and O. Boucher, *Rev. Geophys.*, 2000, **38**, 513–543.

[41] M. Kulmala, J. Kontkanen, H. Junninen, K. Lehtipalo, H. E. Manninen, T. Nieminen, T. Petäjä, M. Sipilä, S. Schobesberger, P. Rantala, A. Franchin, T. Jokinen, E. Järvinen, M. Äijälä, J. Kangasluoma, J. Hakala, P. P. Aalto, P. Paasonen, J. Mikkilä, J. Vanhanen, J. Aalto, H. Hakola, U. Makkonen, T. Ruuskanen, R. L. Mauldin, J. Duplissy, H. Vehkamäki, J. Bäck, A. Kortelainen, I. Riipinen, T. Kurtén, M. V. Johnston, J. N. Smith, M. Ehn, T. F. Mentel, K. E. J. Lehtinen, A. Laaksonen, V.-M. Kerminen and D. R. Worsnop, *Science*, 2013, **339**, 943–946.

[42] M. Kulmala, A. Toivonen, J. M. Mäkelä and A. Laaksonen, *Tellus B: Chem. Phys. Meteorol.*, 1998, **50**, 449–462.

[43] M. Sipilä, T. Berndt, T. Petäjä, D. Brus, J. Vanhanen, F. Stratmann, J. Patokoski, R. L. Mauldin, A.-P. Hyvärinen, H. Lihavainen *et al.*, *Science*, 2010, **327**, 1243–1246.

[44] A. Metzger, B. Verheggen, J. Dommen, J. Duplissy, A. S. Prevot, E. Weingartner, I. Riipinen, M. Kulmala, D. V. Spracklen, K. S. Carslaw *et al.*, *Proc. Natl. Acad. Sci.*, 2010, **107**, 6646–6651.

[45] D. Wimmer, K. Lehtipalo, A. Franchin, J. Kangasluoma, F. Kreissl, A. Kürten, A. Kupc, A. Metzger, J. Mikkilä, T. Petäjä *et al.*, *Atmos. Meas. Tech.*, 2013, **6**, 1793.

[46] K. Iida, M. R. Stolzenburg and P. H. McMurry, *Aerosol Sci. Technol.*, 2009, **43**, 81–96.

[47] M. Sipilä, K. Lehtipalo, M. Kulmala, T. Petäjä, H. Junninen, P. Aalto, H. Manninen, E.-M. Kyrö, E. Asmi, I. Riipinen *et al.*, *Atmos. Chem. Phys.*, 2008, **8**, 4049–4060.

[48] T. Jokinen, M. Sipilä, H. Junninen, M. Ehn, G. Lönn, J. Hakala, T. Petäjä, R. L. Mauldin III, M. Kulmala and D. R. Worsnop, *Atmos. Chem. Phys.*, 2012, **12**, 4117–4125.

[49] H. Junninen, M. Ehn, T. Petäjä, L. Luosujärvi, T. Kotiaho, R. Kostiainen, U. Rohner, M. Gonin, K. Fuhrer, M. Kulmala and D. R. Worsnop, *Atmos. Meas. Tech.*, 2010, **3**, 1039–1053.

[50] M. Kulmala, I. Riipinen, M. Sipilä, H. E. Manninen, T. Petäjä, H. Junninen, M. D. Maso, G. Mordas, A. Mirme, M. Vana, A. Hirsikko, L. Laakso, R. M. Harrison, I. Hanson, C. Leung, K. E. J. Lehtinen and V.-M. Kerminen, *Science*, 2007, **318**, 89–92.

[51] J. Kazil, P. Stier, K. Zhang, J. Quaas, S. Kinne, D. O'Donnell, S. Rast, M. Esch, S. Ferrachat, U. Lohmann and J. Feichter, *Atmos. Chem. Phys. Discuss.*, 2010, **10**, 12261–12308.

[52] R. Zhang, A. Khalizov, L. Wang, M. Hu and W. Xu, *Chem. Rev.*, 2012, **112**, 1957–2011.

[53] M. Kulmala, *Science*, 2003, **302**, 1000–1001.

[54] S. J. Smith, H. Pitcher and T. M. Wigley, *Glob. Planet. Change*, 2001, **29**, 99–119.

[55] S. J. Smith, J. v. Aardenne, Z. Klimont, R. J. Andres, A. Volke and S. Delgado Arias, *Atmos. Chem. Phys.*, 2011, **11**, 1101–1116.

[56] G. J. Doyle, *J. Chem. Phys.*, 1961, **35**, 795–799.

[57] R. J. Weber, P. H. McMurry, F. L. Eisele and D. J. Tanner, *J. Atmos. Sci.*, 1995, **52**, 2242–2257.

[58] V.-M. Kerminen, T. Petäjä, H. E. Manninen, P. Paasonen, T. Nieminen, M. Sipilä, H. Junninen, M. Ehn, S. Gagné, L. Laakso, I. Riipinen, H. Vehkamäki, T. Kurtén, I. K. Ortega, M. Dal Maso, D. Brus, A. Hyvärinen, H. Lihavainen, J. Leppä, K. E. J. Lehtinen, A. Mirme, S. Mirme, U. Hõrrak, T. Berndt, F. Stratmann, W. Birmili, A. Wiedensohler, A. Metzger, J. Dommen, U. Baltensperger, A. Kiendler-Scharr, T. F. Mentel, J. Wildt, P. M. Winkler, P. E. Wagner, A. Petzold, A. Minikin, C. Plass-Dülmer, U. Pöschl, A. Laaksonen and M. Kulmala, *Atmos. Chem. Phys.*, 2010, **10**, 10829–10848.

[59] C. Kuang, P. H. McMurry, A. V. McCormick and F. L. Eisele, *J. Geophys. Res. Atmos.*, 2008, **113**, D10209.

[60] J. Almeida, S. Schobesberger, A. Kürten, I. Ortega, O. Kupiainen-Määttä, A. Praplan, A. Adamov, A. Amorim, F. Bianchi, M. Breitenlechner, A. David, J. Dommen, N. Donahue, A. Downard, E. Dunne, J. Duplissy, S. Ehrhart, R. Flagan, A. Franchin, R. Guida, J. Hakala, A. Hansel, M. Heinritzi, H. Henschel, T. Jokinen, H. Junninen, M. Kajos, J. Kangasluoma, H. Keskinen, A. Kupc, T. Kurtén, A. Kvashin, A. Laaksonen, K. Lehtipalo, M. Leiminger, J. Leppä, V. Loukonen, V. Makhmutov, S. Mathot, M. McGrath, T. Nieminen, T. Olenius, A. Onnela, T. Petäjä, F. Riccobono, I. Riipinen, M. Rissanen, L. Rondo, T. Ruuskanen, F. Santos, N. Sarnela, S. Schallhart, R. Schnitzhofer, J. Seinfeld, M. Simon, M. Sipilä, Y. Stozhkov, F. Stratmann, A. Tomé, J. Tröstl, G. Tsagkogeorgas, P. Vaattovaara, Y. Viisanen, A. Virtanen, A. Vrtala, P. Wagner, E. Weingartner, H. Wex, C. Williamson, D. Wimmer, P. Ye, T. Yli-Juuti, K. Carslaw, M. Kulmala, J. Curtius, U. Baltensperger, D. Worsnop, H. Vehkamäki and J. Kirkby, *Nature*, 2013, **502**, 359–363.

[61] K. E. Anderson, J. I. Siepmann, P. H. McMurry and J. VandeVondele, *J. Am. Chem. Soc.*, 2008, **130**, 14144–14147.

[62] H. Vehkamäki, M. Kulmala, I. Napari, K. Lehtinen, C. Timmreck, M. Noppel and A. Laaksonen, *J. Geophys. Res. Atmos.*, 2013, **118**, 9330–9330.

[63] R. Weber, J. Marti, P. McMurry, F. Eisele, D. Tanner and A. Jefferson, *Chem. Eng. Commun.*, 1996, **151**, 53–64.

[64] H. Henschel, T. Kurtén and H. Vehkamäki, *J. Phys. Chem. A*, 2016, **120**, 1886–1896.

[65] G. K. Yue and P. Hamill, *J. Aerosol Sci.*, 1979, **10**, 609–614.

[66] P. Mirabel and J. L. Katz, *J. Chem. Phys.*, 1974, **60**, 1138–1144.

[67] A. Jaecker-Voirol and P. Mirabel, *Atmos. Environ. (1967)*, 1989, **23**, 2053–2057.

[68] T. Petäjä, M. Sipilä, P. Paasonen, T. Nieminen, T. Kurtén, I. K. Ortega, F. Stratmann, H. Vehkamäki, T. Berndt and M. Kulmala, *Phys. Rev. Lett.*, 2011, **106**, 228302.

[69] A. Bouwman, D. Lee, W. Asman, F. Dentener, K. Van Der Hoek and J. Olivier, *Global biogeochem. cycles*, 1997, **11**, 561–587.

[70] M. Sutton, U. Dragosits, Y. Tang and D. Fowler, *Atmos. Environ.*, 2000, **34**, 855–869.

[71] N. Anderson, R. Strader and C. Davidson, *Environ. Int.*, 2003, **29**, 277–286.

[72] E. Buijsman, H. F. Maas and W. A. Asman, *Atmos. Environ. (1967)*, 1987, **21**, 1009–1022.

[73] D. J. Coffman and D. A. Hegg, *J. Geophys. Res.*, 1995, **100**, 7147–7160.

[74] P. Korhonen, M. Kulmala, A. Laaksonen, Y. Viisanen, R. McGraw and J. Seinfeld, *J. Geophys. Res. Atmos.*, 1999, **104**, 26349–26353.

[75] T. Kurtén, L. Torpo, C.-G. Ding, H. Vehkamäki, M. R. Sundberg, K. Laasonen and M. Kulmala, *J. Geophys. Res. Atmos.*, 2007, **112**, D04210.

[76] T. Kurtén, L. Torpo, M. R. Sundberg, V.-M. Kerminen, H. Vehkamäki and M. Kulmala, *Atmos. Chem. Phys.*, 2007, **7**, 2765–2773.

[77] S. M. Ball, D. R. Hanson, F. L. Eisele and P. H. McMurry, *J. Geophys. Res.*, 1999, **104**, 23709–23718.

[78] W.-W. Liu, X.-L. Wang, S.-L. Chen, Y.-H. Zhang and Z.-S. Li, *Theor. Chem. Acc.*, 2012, **131**, 1103.

[79] J. Kirkby, J. Curtius, J. Almeida, E. Dunne, J. Duplissy, S. Ehrhart, A. Franchin, S. Gagné, L. Ickes, A. Kürten *et al.*, *Nature*, 2011, **476**, 429–433.

[80] K.-L. Ho, Y.-C. Chung, Y.-H. Lin and C.-P. Tseng, *Chemosphere*, 2008, **72**, 250–256.

[81] X. Ge, A. S. Wexler and S. L. Clegg, *Atmos. Environ.*, 2011, **45**, 524–546.

[82] G. W. Schade and P. J. Crutzen, *J. Atmos. Chem.*, 1995, **22**, 319–346.

[83] G. Hutchinson, A. Mosier and C. Andre, *J. Environ. Qual.*, 1982, **11**, 288–293.

[84] T. Kurtén, V. Loukonen, H. Vehkamäki and M. Kulmala, *Atmos. Chem. Phys.*, 2008, **8**, 4095–4103.

[85] N. Freshour, K. Carlson, Y. Melka, S. Hinz, B. Panta and D. Hanson, *Atmos. Meas. Tech.*, 2014, **7**, 3611.

[86] M. Erupe, A. Viggiano and S.-H. Lee, *Atmos. Chem. Phys.*, 2011, **11**, 4767–4775.

[87] R. Atkinson, R. Perry and J. Pitts Jr, *J. Chem. Phys.*, 1978, **68**, 1850–1853.

[88] V. Loukonen, T. Kurtén, I. K. Ortega, H. Vehkamäki, A. A. H. Padua, K. Sellegri and M. Kulmala, *Atmos. Chem. Phys.*, 2010, **10**, 4961–4974.

[89] A. B. Nadykto, F. Yu, M. V. Jakovleva, J. Herb and Y. Xu, *Entropy*, 2011, **13**, 554–569.

[90] J. Elm, C. N. Jen, T. Kurtén and H. Vehkamäki, *J. Phys. Chem. A*, 2016, **120**, 3693–3700.

[91] C. N. Jen, R. Bachman, J. Zhao, P. H. McMurry and D. R. Hanson, *Geophys. Res. Lett.*, 2016, **43**, 867–873.

[92] J. Elm, M. Passananti, T. Kurtén and H. Vehkamäki, *J. Phys. Chem. A*, 2017, **121**, 6155–6164.

[93] F. Raes, A. Janssens and R. Van Dingenen, *J. Aerosol Sci.*, 1986, **17**, 466–470.

[94] F. Yu and R. P. Turco, *Geophys. Res. Lett.*, 2000, **27**, 883–886.

[95] E. R. Lovejoy, J. Curtius and K. D. Froyd, *J. Geophys. Res.*, 2004, **109**, D08204.

[96] I. Ortega, T. Olenius, O. Kupiainen-Määttä, V. Loukonen, T. Kurtén and H. Vehkamäki, *Atmos. Chem. Phys.*, 2014, **14**, 7995–8007.

[97] J. Kirkby, J. Duplissy, K. Sengupta, C. Frege, H. Gordon, C. Williamson, M. Heinritzi, M. Simon, C. Yan, J. Almeida, J. Trostl, T. Nieminen, I. Ortega, R. Wagner, A. Adamov, A. Amorim, A.-K. Bernhammer, F. Bianchi, M. Breitenlechner, S. Brilke, X. Chen, J. Craven, A. Dias, S. Ehrhart, R. Flagan, A. Franchin, C. Fuchs, R. Guida, J. Hakala, C. Hoyle, T. Jokinen, H. Junninen, J. Kangasluoma, J. Kim, M. Krapf, A. Kurtén, A. Laaksonen, K. Lehtipalo, V. Makhmutov, S. Mathot, U. Molteni, A. Onnela, O. Peräkyla, F. Piel, T. Petaja, A. Praplan, K. Pringle, A. Rap, N. Richards, I. Riipinen, M. Rissanen, L. Rondo, N. Sarnela, S. Schobesberger, C. Scott, J. Seinfeld, M. Sipilä, G. Steiner, Y. Stozhkov, F. Stratmann, A. Tomé, A. Virtanen, A. Vogel, A. Wagner,

P. Wagner, E. Weingartner, D. Wimmer, P. Winkler, P. Ye, X. Zhang, A. Hansel, J. Dommen, N. Donahue, D. Worsnop, U. Baltensperger, M. Kulmala, K. Carslaw and J. Curtius, *Nature*, 2016, **533**, 521–526.

[98] V. A. Mohnen, *J. Geophys. Res.*, 1970, **75**, 1717–1721.

[99] H. Svensmark and E. Friis-Christensen, *J. Atmos. Sol.-Terr. Phys.*, 1997, **59**, 1225–1232.

[100] F. Yu and R. P. Turco, *J. Geophys. Res. Atmos.*, 2001, **106**, 4797–4814.

[101] M. Wilkening, *J. Geophys. Res. Atmos.*, 1985, **90**, 5933–5935.

[102] K. S. Carslaw, R. G. Harrison and J. Kirkby, *Science*, 2002, **298**, 1732–1737.

[103] H. Manninen, T. Nieminen, E. Asmi, S. Gagné, S. Häkkinen, K. Lehtipalo, P. Aalto, M. Vana, A. Mirme, S. Mirme *et al.*, *Atmos. Chem. Phys.*, 2010, **10**, 7907–7927.

[104] J. Kirkby, J. Curtius, J. Almeida, E. Dunne, J. Duplissy, S. Ehrhart, A. Franchin, S. Gagné, L. Ickes, A. Kürten, A. Kupc, A. Metzger, F. Riccobono, L. Rondo, S. Schobesberger, G. Tsagkogeorgas, D. Wimmer, A. Amorim, F. Bianchi, M. Breitenlechner, A. David, J. Dommen, A. Downard, M. Ehn, R. Flagan, S. Haider, A. Hansel, D. Hauser, W. Jud, H. Junninen, F. Kreissl, A. Kvashin, A. Laaksonen, K. Lehtipalo, J. Lima, E. Lovejoy, V. Makhmutov, S. Mathot, J. Mikkilä, P. Minginette, S. Mogo, T. Nieminen, A. Onnela, P. Pereira, T. Petäjä, R. Schnitzhofer, J. Seinfeld, M. Sipilä, Y. Stozhkov, F. Stratmann, A. Tomé, J. Vanhanen, Y. Viisanen, A. Virtala, P. Wagner, H. Walther, E. Weingartner, H. Wex, P. Winkler, K. Carslaw, D. Worsnop, U. Baltensperger and K. M., *Nature*, 2013, **476**, 429–433.

[105] S. Gagné, T. Nieminen, T. Kurtén, H. Manninen, T. Petäjä, L. Laakso, V.-M. Kerminen, M. Boy and M. Kulmala, *Atmos. Chem. Phys.*, 2010, **10**, 3743–3757.

[106] D. E. Husar, B. Temelso, A. L. Ashworth and G. C. Shields, *J. Phys. Chem. A*, 2012, **116**, 5151–5163.

[107] J. Slowik, C. Stroud, J. Bottenheim, P. Brickell, R.-W. Chang, J. Liggio, P. Makar, R. Martin, M. Moran, N. Shantz *et al.*, *Atmos. Chem. Phys.*, 2010, **10**, 2825–2845.

[108] A. B. Nadykto and F. Yu, *Chem. Phys. Lett.*, 2007, **435**, 14–18.

[109] A. Guenther, C. N. Hewitt, D. Erickson, R. Fall, C. Geron, T. Graedel, P. Harley, L. Klinger, M. Lerdau, W. McKay *et al.*, *J. Geophys. Res. Atmos.*, 1995, **100**, 8873–8892.

[110] J. Kesselmeier and M. Staudt, *J. Atmos. Chem.*, 1999, **33**, 23–88.

[111] S. D. Piccot, J. J. Watson and J. W. Jones, *J. Geophys. Res. Atmos.*, 1992, **97**, 9897–9912.

[112] J. H. Seinfeld and J. F. Pankow, *Annu. Rev. Phys. Chem.*, 2003, **54**, 121–140.

[113] J. Bäck, J. Aalto, M. Henriksson, H. Hakola, Q. He and M. Boy, *Biogeosciences*, 2012, **9**, 689–702.

[114] A. B. Guenther, P. R. Zimmerman, P. C. Harley, R. K. Monson and R. Fall, *J. Geophys. Res. Atmos.*, 1993, **98**, 12609–12617.

[115] T. Berndt, S. Richters, R. Kaethner, J. Voigtländer, F. Stratmann, M. Sipilä, M. Kulmala and H. Herrmann, *J. Phys. Chem. A*, 2015, **119**, 10336–10348.

[116] R. Mauldin Iii, T. Berndt, M. Sipilä, P. Paasonen, T. Petäjä, S. Kim, T. Kurtén, F. Stratmann, V.-M. Kerminen and M. Kulmala, *Nature*, 2012, **488**, 193–196.

[117] I. Riipinen, T. Yli-Juuti, J. R. Pierce, T. Petäjä, D. R. Worsnop, M. Kulmala and N. M. Donahue, *Nat. Geosci.*, 2012, **5**, 453.

[118] X. Zhang, R. C. McVay, D. D. Huang, N. F. Dalleska, B. Aumont, R. C. Flagan and J. H. Seinfeld, *Proc. Natl. Acad. Sci.*, 2015, **112**, 14168–14173.

[119] T. Hoffmann, R. Bandur, U. Marggraf and M. Linscheid, *J. Geophys. Res. Atmos.*, 1998, **103**, 25569–25578.

[120] T. Christoffersen, J. Hjorth, O. Horie, N. Jensen, D. Kotzias, L. Molander, P. Neeb, L. Ruppert, R. Winterhalter, A. Virkkula, K. Wirtz and B. Larsen, *Atmos. Environ.*, 1998, **32**, 1657–1661.

[121] M. Hallquist, J. Wenger, U. Baltensperger, Y. Rudich, D. Simpson, M. Claeys, J. Dommen, N. Donahue, C. George, A. Goldstein *et al.*, *Atmos. Chem. Phys.*, 2009, **9**, 5155–5236.

[122] J. H. Kroll and J. H. Seinfeld, *Atmos. Environ.*, 2008, **42**, 3593–3624.

[123] T. Yli-Juuti, A. A. Zardini, A. C. Eriksson, A. M. K. Hansen, J. H. Pagels, E. Swietlicki, B. Svenningsson, M. Glasius, D. R. Worsnop, I. Riipinen and M. Bilde, *Environ. Sci. Technol.*, 2013, **47**, 12123–12130.

[124] S. A. K. Häkkinen, V. F. McNeill and I. Riipinen, *Environ. Sci. Technol.*, 2014, **48**, 13718–13726.

[125] S. Schobesberger, H. Junninen, F. Bianchi, G. Lönn, M. Ehn, K. Lehtipalo, J. Dommen, S. Ehrhart, I. K. Ortega, A. Franchin, T. Nieminen, F. Riccobono, M. Hutterli, J. Duplissy, J. Almeida, A. Amorim, M. Breitenlechner, A. J. Downard, E. M. Dunne, R. C. Flagan, M. Kajos, H. Keskinen, J. Kirkby, A. Kupc, A. Kürten, T. Kurtén, A. Laaksonen, S. Mathot, A. Onnela, A. P. Praplan, L. Rondo, F. D. Santos, S. Schallhart, R. Schnitzhofer, M. Sipilä, A. Tomé, G. Tsagkogeorgas, H. Vehkamäki, D. Wimmer, U. Baltensperger, K. S. Carslaw, J. Curtius, A. Hansel, T. Petäjä, M. Kulmala, N. M. Donahue and D. R. Worsnop, *Proc. Natl. Acad. Sci.*, 2013, **110**, 17223–17228.

[126] M. Ehn, J. A. Thornton, E. Kleist, M. Sipilä, H. Junninen, I. Pullinen, M. Springer, F. Rubach, R. Tillmann, B. Lee, F. Lopez-Hilfiker, S. Andres, I.-H. Acir, M. Rissanen, T. Jokinen, S. Schobesberger, J. Kangasluoma, J. Kontkanen, T. Nieminen, T. Kurtén, L. B. Nielsen, S. Jørgensen, H. G. Kjærgaard, M. Canagaratna, M. Dal Maso, T. Berndt, T. Petaejae, A. Wahner, V.-M. Kerminen, M. Kulmala, D. R. Worsnop, J. Wildt and T. F. Mentel, *Nature*, 2014, **506**, 476–479.

[127] A. L. Paciga, I. Riipinen and S. N. Pandis, *Environ. Sci. Technol.*, 2014, **48**, 13769–13775.

[128] F. Riccobono, S. Schobesberger, C. E. Scott, J. Dommen, I. K. Ortega, L. Rondo, J. Almeida, A. Amorim, F. Bianchi, M. Breitenlechner, A. David, A. Downard, E. M. Dunne, J. Duplissy, S. Ehrhart, R. C. Flagan, A. Franchin, A. Hansel, H. Junninen, M. Kajos, H. Keskinen, A. Kupc, A. Kürten, A. N. Kvashin, A. Laaksonen, K. Lehtipalo, V. Makhmutov, S. Mathot, T. Nieminen, A. Onnela, T. Petäjä, A. P. Praplan, F. D. Santos, S. Schallhart, J. H. Seinfeld, M. Sipilä, D. V. Spracklen, Y. Stozhkov, F. Stratmann, A. Tomé, G. Tsagkogeorgas, P. Vaattovaara, Y. Viisanen, A. Vrtala, P. E. Wagner, E. Weingartner, H. Wex, D. Wimmer, K. S. Carslaw, J. Curtius, N. M. Donahue, J. Kirkby, M. Kulmala, D. R. Worsnop and U. Baltensperger, *Science*, 2014, **344**, 717–721.

[129] M. P. Rissanen, T. Kurtén, M. Sipilä, J. A. Thornton, J. Kangasluoma, N. Sarnela, H. Junninen, S. Jørgensen, S. Schallhart, M. K. Kajos *et al.*, *J. Am. Chem. Soc.*, 2014, **136**, 15596–15606.

[130] J. D. Crounse, L. B. Nielsen, S. Jørgensen, H. G. Kjærgaard and P. O. Wennberg, *J. Phys. Chem. Lett.*, 2013, **4**, 3513–3520.

[131] T. Jokinen, M. Sipilä, S. Richters, V.-M. Kerminen, P. Paasonen, F. Stratmann, D. Worsnop, M. Kulmala, M. Ehn, H. Herrmann and T. Berndt, *Angew. Chem. Int. Ed.*, 2014, **53**, 14596–14600.

[132] T. Kurtén, K. Tiusanen, P. Roldin, M. Rissanen, J.-N. Luy, M. Boy, M. Ehn and N. Donahue, *J. Phys. Chem. A*, 2016, **120**, 2569–2582.

[133] T. Kurtén, M. P. Rissanen, K. Mackeprang, J. A. Thornton, N. Hyttinen, S. Jørgensen, M. Ehn and H. G. Kjærgaard, *J. Phys. Chem. A*, 2015, **119**, 11366–11375.

[134] M. Ehn, J. A. Thornton, E. Kleist, M. Sipilä, H. Junninen, I. Pullinen, M. Springer, F. Rubach, R. Tillmann, B. Lee *et al.*, *Nature*, 2014, **506**, 476.

[135] R. Szmigielski, J. D. Surratt, Y. Gómez-González, P. Van der Veken, I. Kourtchev, R. Vermeylen, F. Blockhuys, M. Jaoui, T. E. Kleindienst, M. Lewandowski, J. H. Offenberg, E. O. Edney, J. H. Seinfeld, W. Maenhaut and M. Claeys, *Geophys. Res. Lett.*, 2007, **34**, 24811.

[136] T. Hoffmann, C. D. O'Dowd and J. H. Seinfeld, *Geophys. Res. Lett.*, 2001, **28**, 1949–1952.

[137] B. Alicke, K. Hebestreit, J. Stutz and U. Platt, *Nature*, 1999, **397**, 572.

[138] M. Sipilä, N. Sarnela, T. Jokinen, H. Henschel, H. Junninen, J. Kontkanen, S. Richters, J. Kangasluoma, A. Franchin, O. Peräkylä *et al.*, *Nature*, 2016, **537**, 532.

[139] N. Bork, J. Elm, T. Olenius and H. Vehkamäki, *Atmos. Chem. Phys.*, 2014, **14**, 12023–12030.

[140] P. Atkins and J. De Paula, *Atkins' Physical Chemistry*, Oxford University Press, 8th edn., 2006.

[141] H. Vehkamäki and I. Riipinen, *Chem. Soc. Rev.*, 2012, **41**, 5160–5173.

[142] H. Reiss, *J. Chem. Phys.*, 1950, **18**, 840–848.

[143] I. K. Ortega, O. Kupiainen, T. Kurtén, T. Olenius, O. Wilkman, M. J. McGrath, V. Loukonen and H. Vehkamäki, *Atmos. Chem. Phys.*, 2012, **12**, 225–235.

[144] B. R. Bzdek, J. W. DePalma, D. P. Ridge, J. Laskin and M. V. Johnston, *J. Am. Chem. Soc.*, 2013, **135**, 3276–3285.

[145] J. W. DePalma, B. R. Bzdek, D. P. Ridge and M. V. Johnston, *J. Phys. Chem. A*, 2014, **118**, 11547–11554.

[146] D. W. Oxtoby, *Acc. Chem. Res.*, 1998, **31**, 91–97.

[147] F. F. Abraham, *Homogeneous Nucleation Theory - The Pretransition Theory of Vapor Condensation*, Academic Press, New York, 1974.

[148] H. Vehkamäki, *Classical nucleation theory in multicomponent systems*, Springer Science & Business Media, 2006.

[149] P. H. McMurry, *J. Colloid. Interface. Sci.*, 1980, **78**, 513–527.

[150] T. Olenius, O. Kupiainen, I. K. Ortega and H. Vehkamäki, *AIP Conf. Proc.*, 2013, **1527**, 212–215.

[151] T. Olenius, O. Kupiainen-Määttä, I. Ortega, T. Kurtén and H. Vehkamäki, *J. Chem. Phys.*, 2013, **139**, 084312.

[152] D. Kashchiev, *J. Chem. Phys.*, 1982, **76**, 5098–5102.

[153] H. Vehkamäki, M. J. McGrath, T. Kurtén, J. Julin, K. E. Lehtinen and M. Kulmala, *J. Chem. Phys.*, 2012, **136**, 094107.

[154] R. McGraw and D. T. Wu, *J. Chem. Phys.*, 2003, **118**, 9337–9347.

[155] D. Kashchiev, *Nucleation - Basic Theory With Applications*, Butterworth-Heinemann, Oxford, 2000.

[156] O. Kupiainen-Määttä, T. Olenius, H. Korhonen, J. Malila, M. Dal Maso, K. Lehtinen and H. Vehkamäki, *J. Aerosol Sci.*, 2014, **77**, 127–144.

[157] C. Rey, L. Gallego, M. Iniguez and J. Alonso, *Physica B: Condens. Matter*, 1992, **179**, 273–277.

[158] C. Roman and I. Garzon, *Z. Phys. D Atoms, Molecules and Clusters*, 1991, **20**, 163–166.

[159] V. Loukonen, N. Bork and H. Vehkamäki, *Mol. Phys.*, 2014, **112**, 1979–1986.

[160] S. Chapman and T. G. Cowling, *The mathematical theory of non-uniform gases*, University Press, Cambridge, 1970.

[161] O. Kupiainen-Määttä, T. Olenius, T. Kurtén and H. Vehkamäki, *J. Phys. Chem. A*, 2013, **117**, 14109–14119.

[162] T. Su and W. J. Chesnavich, *J. Chem. Phys.*, 1982, **76**, 5183–5185.

[163] W. J. Chesnavich, T. Su and M. T. Bowers, *J. Chem. Phys.*, 1980, **72**, 2641–2655.

[164] M. J. McGrath, T. Olenius, I. K. Ortega, V. Loukonen, P. Paasonen, T. Kurtén, M. Kulmala and H. Vehkamäki, *Atmos. Chem. Phys.*, 2012, **12**, 2345–2355.

[165] L. Shampine and M. Reichelt, *SIAM J. Sci. Comput.*, 1997, **18**, 1–22.

[166] L. Farkas, *Z. Phys. Chem.*, 1927, **125**, 236–242.

[167] R. Becker and W. Döring, *Ann. Phys.*, 1935, **416**, 719–752.

[168] N. Allinger, *J. Am. Chem. Soc.*, 1977, **99**, 8127–8134.

[169] N. Allinger, *Adv. Phys. Org. Chem.*, 1976, **13**, 1–82.

[170] T. Helgaker, P. Jorgensen and J. Olsen, *Molecular electronic-structure theory*, John Wiley & Sons, 2014.

[171] C. Jacoboni and L. Reggiani, *Rev. Mod. Phys.*, 1983, **55**, 645.

[172] R. Car and M. Parrinello, *Phys. Rev. Lett.*, 1985, **55**, 2471.

[173] W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell and P. A. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 5179–5197.

[174] S. Lifson and A. Warshel, *J. Chem. Phys.*, 1968, **49**, 5116–5129.

[175] T. Olenius, I. Riipinen, K. Lehtipalo and H. Vehkamäki, *J. Aerosol Sci.*, 2014, **78**, 55–70.

[176] J. Merikanto, E. Zapadinsky, A. Lauri and H. Vehkamäki, *Phys. Rev. Lett.*, 2007, **98**, 145702.

[177] S. Kirkpatrick, C. D. Gelatt and M. P. Vecchi, *Science*, 1983, **220**, 671–680.

[178] R. G. Parr, *Density functional theory of atoms and molecules*, Springer, 1980.

[179] J. Elm, M. Bilde and K. V. Mikkelsen, *J. Phys. Chem. A*, 2013, **117**, 6695–6701.

[180] J. Elm, M. Fard, M. Bilde and K. V. Mikkelsen, *J. Phys. Chem. A*, 2013, **117**, 12990–12997.

[181] J. J. Stewart, *J. Mol. Model.*, 2007, **13**, 1173–1213.

[182] B. Temelso, T. E. Morrell, R. M. Shields, M. A. Allodi, E. K. Wood, K. N. Kirschner, T. C. Castonguay, K. A. Archer and G. C. Shields, *J. Phys. Chem. A*, 2012, **116**, 2209–2224.

[183] L. Partanen, H. Vehkamäki, K. Hansen, J. Elm, H. Henschel, T. Kurtén, R. Halonen and E. Zapadinsky, *J. Phys. Chem. A*, 2016, **120**, 8613–8624.

[184] P. A. M. Dirac, Proceedings of the Royal Society of London A: Mathematical, Physical and Engineering Sciences, 1929, pp. 714–733.

[185] E. Schrödinger, *Statistical Thermodynamics*, Dover Publications, 1989.

[186] F. Jensen, *Introduction to Computational Chemistry*, John Wiley and Sons, 2nd edn., 2007.

[187] M. Born and R. Oppenheimer, *Ann. Phys.*, 1927, **389**, 457–484.

[188] P. Atkins and R. Friedman, *Molecular Quantum Mechanics*, Oxford University Press, 4th edn., 2005.

[189] P. Bunker, *J. Mol. Spec.*, 1972, **42**, 478–494.

[190] P. Bunker and R. Moss, *Mol. Phys.*, 1977, **33**, 417–424.

[191] L. Che, Z. Ren, X. Wang, W. Dong, D. Dai, X. Wang, D. H. Zhang, X. Yang, L. Sheng, G. Li *et al.*, *Science*, 2007, **317**, 1061–1064.

[192] E. Schrödinger, Mathematical Proceedings of the Cambridge Philosophical Society, 1935, pp. 555–563.

[193] J. H. Shirley, *Phys. Rev.*, 1965, **138**, B979.

[194] E. Balslev and J.-M. Combes, *Commun. Mat. Phys.*, 1971, **22**, 280–294.

[195] J. Snijders, E. Baerends and P. Ros, *Mol. Phys.*, 1979, **38**, 1909–1929.

[196] D. J. Gorin and F. D. Toste, *Nature*, 2007, **446**, 395.

[197] P. Pyykkö, *Angew. Chem. Int. Ed.*, 2004, **43**, 4412–4456.

[198] L. J. Norrby, *J. Chem. Educ*, 1991, **68**, 110.

[199] C. J. Cramer, *Essentials of Computational Chemistry - Theories and Models*, John Wiley and Sons, 2nd edn., 2004.

[200] S. Huzinaga, D. McWilliams and B. Domsky, *J. Chem. Phys.*, 1971, **54**, 2283–2284.

[201] S. Huzinaga, *J. Chem. Phys.*, 1965, **42**, 1293–1302.

[202] W. J. Hehre, R. F. Stewart and J. A. Pople, *J. Chem. Phys.*, 1969, **51**, 2657–2664.

[203] R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650–654.

[204] M. M. Francl, W. J. Pietro, W. J. Hehre, J. S. Binkley, M. S. Gordon, D. J. DeFrees and J. A. Pople, *J. Chem. Phys.*, 1982, **77**, 3654–3665.

[205] T. Clark, J. Chandrasekhar, G. W. Spitznagel and P. V. R. Schleyer, *J. Comp. Chem.*, 1983, **4**, 294–301.

[206] E. Papajak, J. Zheng, X. Xu, H. R. Leverentz and D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 3027–3034.

[207] M. Gutowski, J. Van Lenthe, J. Verbeek, F. Van Duijneveldt and G. Chałasinski, *Chem. Phys. Lett.*, 1986, **124**, 370–375.

[208] S. Simon, M. Duran and J. Dannenberg, *J. Chem. Phys.*, 1996, **105**, 11024–11031.

[209] S. F. Boys and F. d. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.

[210] T. H. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007–1023.

[211] D. E. Woon and T. H. Dunning Jr, *J. Chem. Phys.*, 1993, **98**, 1358–1371.

[212] R. A. Kendall, T. H. Dunning and R. J. Harrison, *J. Chem. Phys.*, 1992, **96**, 6796–6806.

[213] J. C. Slater, *Phys. Rev.*, 1951, **81**, 385.

[214] R. Boyd and C. Coulson, *J. Phys. B: At. Mol. Opt. Phys.*, 1974, **7**, 1805.

[215] D. Bohm and D. Pines, *Phys. Rev.*, 1953, **92**, 609.

[216] E. Baerends, D. Ellis and P. Ros, *Chem. Phys.*, 1973, **2**, 41–51.

[217] R. M. Martin, *Electronic Structure: Basic Theory and Practical Methods*, Cambridge University Press, 2005.

[218] J. Olsen, P. Jørgensen and J. Simons, *Chem. Phys. Lett.*, 1990, **169**, 463–472.

[219] U. Von Barth and C. Gelatt, *Phys. Rev. B*, 1980, **21**, 2222.

[220] H.-J. Werner and P. J. Knowles, *J. Chem. Phys.*, 1985, **82**, 5053–5063.

[221] B. O. Roos, *Adv. Chem. Phys.: Ab Initio Methods in Quantum Chemistry Part 2*, 2007, **69**, 399–445.

[222] J. Finley, P.-Å. Malmqvist, B. O. Roos and L. Serrano-Andrés, *Chem. Phys. Lett.*, 1998, **288**, 299–306.

[223] U. Schollwöck, *Rev. Mod. Phys.*, 2005, **77**, 259.

[224] S. R. White, *Phys. Rev. Lett.*, 1992, **69**, 2863.

[225] J. A. Pople, M. Head-Gordon and K. Raghavachari, *J. Chem. Phys.*, 1987, **87**, 5968–5975.

[226] C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618.

[227] R. J. Bartlett and M. Musiał, *Rev. Mod. Phys.*, 2007, **79**, 291.

[228] P. Knowles and N. Handy, *Chem. Phys. Lett.*, 1984, **111**, 315–321.

[229] B. Levy and G. Berthier, *Int. J. Quantum Chem.*, 1968, **2**, 307–319.

[230] R. H. Nobes, J. A. Pople, L. Radom, N. C. Handy and P. J. Knowles, *Chem. Phys. Lett.*, 1987, **138**, 481–485.

[231] M. Head-Gordon, J. A. Pople and M. J. Frisch, *Chem. Phys. Lett.*, 1988, **153**, 503–506.

[232] M. Kállay and P. R. Surján, *J. Chem. Phys.*, 2001, **115**, 2945–2954.

[233] G. D. Purvis III and R. J. Bartlett, *J. Chem. Phys.*, 1982, **76**, 1910–1918.

[234] G. E. Scuseria, C. L. Janssen and H. F. Schaefer Iii, *J. Chem. Phys.*, 1988, **89**, 7382–7387.

[235] O. Christiansen, H. Koch and P. Jørgensen, *Chem. Phys. Lett.*, 1995, **243**, 409–418.

[236] C. Hättig, *J. Chem. Phys.*, 2003, **118**, 7751–7761.

[237] C. Hättig and F. Weigend, *J. Chem. Phys.*, 2000, **113**, 5154–5161.

[238] T. Kurtén, B. Bonn, H. Vehkamäki and M. Kulmala, *J. Phys. Chem. A*, 2007, **111**, 3394–3401.

[239] I. K. Ortega, T. Kurtén, H. Vehkamäki and M. Kulmala, *Atmos. Chem. Phys.*, 2008, **8**, 2859–2867.

[240] H. Koch, O. Christiansen, P. Jo/rgensen, A. M. Sanchez de Merás and T. Helgaker, *J. Chem. Phys.*, 1997, **106**, 1808–1818.

[241] K. Raghavachari, G. W. Trucks, J. A. Pople and M. Head-Gordon, *Chem. Phys. Lett.*, 1989, **157**, 479–483.

[242] J. Rezac and P. Hobza, *J. Chem. Theory Comput.*, 2013, **9**, 2151–2155.

[243] F. Neese, *J. Comp. Chem.*, 2003, **24**, 1740–1747.

[244] M. Feyereisen, G. Fitzgerald and A. Komornicki, *Chem. Phys. Lett.*, 1993, **208**, 359–363.

[245] B. Dunlap, *J. Mol. Struc.*, 2000, **529**, 37–40.

[246] F. Weigend, M. Häser, H. Patzelt and R. Ahlrichs, *Chem. Phys. Lett.*, 1998, **294**, 143–152.

[247] F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2002, **116**, 3175–3183.

[248] A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen and A. K. Wilson, *Chem. Phys. Lett.*, 1998, **286**, 243–252.

[249] S. Ten-no and J. Noga, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 114–125.

[250] W. Kutzelnigg, *Theor. Chem. Acc.*, 1985, **68**, 445–469.

[251] D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45–48.

[252] L. Kong, F. A. Bischoff and E. F. Valeev, *Chem. Rev.*, 2011, **112**, 75–107.

[253] D. Bokhan, S. Ten-No and J. Noga, *Phys. Chem. Chem. Phys.*, 2008, **10**, 3320–3326.

[254] W. Klopper, F. R. Manby, S. Ten-No and E. F. Valeev, *Int. Rev. Phys. Chem.*, 2006, **25**, 427–468.

[255] C. Hättig, D. P. Tew and A. Köhn, *J. Chem. Phys.*, 2010, **132**, 231102.

[256] K. E. Yousaf and K. A. Peterson, *J. Chem. Phys.*, 2008, **129**, 184108.

[257] K. A. Peterson, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2008, **128**, 084102.

[258] K. A. Peterson, M. K. Kesharwani and J. M. L. Martin, *Mol. Phys.*, 2015, **113**, 1551–1558.

[259] G. Knizia, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2009, **130**, 054104.

[260] D. P. Tew, W. Klopper, C. Neiss and C. Hättig, *Phys. Chem. Chem. Phys.*, 2007, **9**, 1921–1930.

[261] J. Zhang and E. F. Valeev, *J. Chem. Theory Comput.*, 2012, **8**, 3175–3186.

[262] E. F. Valeev, *Chem. Phys. Lett.*, 2004, **395**, 190–195.

[263] G. L. Stoychev, A. A. Auer and F. Neese, *J. Chem. Theory Comput.*, 2017, **13**, 554–562.

[264] S. Saebo and P. Pulay, *Ann. Rev. Phys. Chem.*, 1993, **44**, 213–236.

[265] C. Hampel and H.-J. Werner, *J. Chem. Phys.*, 1996, **104**, 6286–6297.

[266] W. Yang, *Phys. Rev. Lett.*, 1991, **66**, 1438–1441.

[267] W. Yang and T. Lee, *J. Chem. Phys.*, 1995, **103**, 5674–5678.

[268] W. Li and S. Li, *J. Chem. Phys.*, 2004, **121**, 6649–6657.

[269] M. Ziółkowski, B. Jansík, T. Kjærgaard and P. Jørgensen, *J. Chem. Phys.*, 2010, **133**, 014107.

[270] K. Kristensen, M. Ziółkowski, B. Jansík, T. Kjærgaard and P. Jørgensen, *J. Chem. Theory Comput.*, 2011, **7**, 1677–1694.

[271] W. Li, P. Piecuch, J. R. Gour and S. Li, *J. Chem. Phys.*, 2009, **131**, 114109.

[272] S. Li, J. Shen, W. Li and Y. Jiang, *J. Chem. Phys.*, 2006, **125**, 074109.

[273] P. Pulay, *Chem. Phys. Lett.*, 1983, **100**, 151–154.

[274] M. Schütz, G. Hetzer and H.-J. Werner, *J. Chem. Phys.*, 1999, **111**, 5691–5705.

[275] M. Schütz and H.-J. Werner, *J. Chem. Phys.*, 2001, **114**, 661–681.

[276] M. Schütz, *J. Chem. Phys.*, 2000, **113**, 9986–10001.

[277] M. Schütz, *J. Chem. Phys.*, 2002, **116**, 8772–8785.

[278] H.-J. Werner, F. R. Manby and P. J. Knowles, *J. Chem. Phys.*, 2003, **118**, 8149–8160.

[279] M. Schütz and F. R. Manby, *Phys. Chem. Chem. Phys.*, 2003, **5**, 3349–3358.

[280] H.-J. Werner and M. Schütz, *J. Chem. Phys.*, 2011, **135**, 144116.

[281] J. Yang, Y. Kurashige, F. R. Manby and G. K. Chan, *J. Chem. Phys.*, 2011, **134**, 044123.

[282] J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz and H.-J. Werner, *J. Chem. Phys.*, 2012, **136**, 144105.

[283] C. Edmiston and M. Krauss, *J. Chem. Phys.*, 1965, **42**, 1119–1120.

[284] C. Edmiston and M. Krauss, *J. Chem. Phys.*, 1966, **45**, 1833–1839.

[285] W. Meyer, *Int. J. Quantum Chem.*, 1971, **5**, 341–348.

[286] W. Meyer, *Theor. Chem. Acc.*, 1974, **35**, 277–292.

[287] R. Ahlrichs and F. Driessler, *Theor. Chem. Acc.*, 1975, **36**, 275–287.

[288] M. Sparta and F. Neese, *Chem. Soc. Rev.*, 2014, **43**, 5032–5041.

[289] F. Neese, F. Wennmohs and A. Hansen, *J. Chem. Phys.*, 2009, **130**, 114108.

[290] F. Neese, A. Hansen and D. G. Liakos, *J. Chem. Phys.*, 2009, **131**, 064103.

[291] C. Riplinger and F. Neese, *J. Chem. Phys.*, 2013, **138**, 034106.

[292] C. Riplinger, B. Sandhoefer, A. Hansen and F. Neese, *J. Chem. Phys.*, 2013, **139**, 134101.

[293] F. Neese, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 73–78.

[294] F. Neese, F. Wennmohs, U. Becker, D. Bykov, D. Ganyushin, A. Hansen, R. Izsak, D. Liakos, C. Kollmar, S. Kossmann *et al.*, *ORCA – An ab initio, DFT and semiempirical SCF-MO package, version 3.0*, Max Planck Institute for Chemical Energy Conversion, Mülheim/Ruhr, Germany, 2013.

[295] M. S. Frank, G. Schmitz and C. Hättig, *Mol. Phys.*, 2017, **115**, 343–356.

[296] Y. Guo, K. Sivalingam, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2016, **144**, 094111.

[297] F. Pavošević, P. Pinski, C. Riplinger, F. Neese and E. F. Valeev, *J. Chem. Phys.*, 2016, **144**, 144109.

[298] P. Pinski, C. Riplinger, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2015, **143**, 034108.

[299] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev and F. Neese, *J. Chem. Phys.*, 2016, **144**, 024109.

[300] F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

[301] D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin and F. Neese, *J. Chem. Theory Comput.*, 2015, **11**, 1525–1539.

[302] D. G. Liakos and F. Neese, *J. Chem. Theory Comput.*, 2015, **11**, 4054–4063.

[303] F. Neese, *Software update: the ORCA program system, version 4.0*, 2017.

[304] E. Hückel, *Z. Phys. A*, 1931, **70**, 204–286.

[305] M. C. Zerner, *Rev. Comp. Ch.*, 1991, **2**, 313–365.

[306] J. A. Pople, D. P. Santry and G. A. Segal, *J. Chem. Phys.*, 1965, **43**, S129–S135.

[307] J. A. Pople and G. A. Segal, *J. Chem. Phys.*, 1965, **43**, S136–S151.

[308] J. Ridley and M. Zerner, *Theor. Chem. Acc.*, 1973, **32**, 111–134.

[309] J. Pople, D. Beveridge and P. Dobosh, *J. Chem. Phys.*, 1967, **47**, 2026–2033.

[310] W. Koch, *Z. Naturforsch. A*, 1993, **48**, 819–828.

[311] M. J. Dewar and W. Thiel, *Theor. Chem. Acc.*, 1977, **46**, 89–104.

[312] P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.

[313] W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, A1133–A1138.

[314] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford University Press, 1989.

[315] W. Kohn, A. D. Becke and R. G. Parr, *J. Phys. Chem.*, 1996, **100**, 12974–12980.

[316] J. P. Perdew and K. Schmidt, AIP Conference Proceedings, 2001, pp. 1–20.

[317] A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098.

[318] A. D. Becke, *J. Chem. Phys.*, 2014, **140**, 18A301.

[319] J. P. Perdew and W. Yue, *Phys. Rev. B*, 1986, **33**, 8800.

[320] J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865.

[321] J. P. Perdew, J. A. Chevary, S. H. Vosko, K. A. Jackson, M. R. Pederson, D. J. Singh and C. Fiolhais, *Phys. Rev. B*, 1992, **46**, 6671.

[322] J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 146401.

[323] J. P. Perdew, J. Tao, V. N. Staroverov and G. E. Scuseria, *J. Chem. Phys.*, 2004, **120**, 6898–6911.

[324] A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

[325] A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 1372–1377.

[326] C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785.

[327] S. Grimme, *J. Chem. Phys.*, 2006, **124**, 034108.

[328] J. Klimeš and A. Michaelides, *J. Chem. Phys.*, 2012, **137**, 120901.

[329] J. P. Perdew and Y. Wang, *Phys. Rev. B*, 1992, **45**, 13244–13249.

[330] Y. Zhao and D. G. Truhlar, *Theor. Chem. Acc.*, 2008, **120**, 215–241.

[331] Y. Zhao and D. G. Truhlar, *J. Phys. Chem. C*, 2008, **112**, 4061–4067.

[332] J. Elm, M. Bilde and K. V. Mikkelsen, *J. Chem. Theory Comput.*, 2012, **8**, 2071–2077.

[333] H. R. Leverentz, J. I. Siepmann, D. G. Truhlar, V. Loukonen and H. Vehkamäki, *J. Phys. Chem. A*, 2013, **117**, 3819–3825.

[334] N. Mardirossian and M. Head-Gordon, *J. Chem. Theory Comput.*, 2013, **9**, 4453–4461.

[335] S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463–1473.

[336] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

[337] S. Grimme, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 211–228.

[338] J.-D. Chai and M. Head-Gordon, *Phys. Chem. Chem. Phys.*, 2008, **10**, 6615–6620.

[339] J. Elm, M. Bilde and K. V. Mikkelsen, *Phys. Chem. Chem. Phys.*, 2013, **15**, 16442–16445.

[340] M. Dion, H. Rydberg, E. Schröder, D. C. Langreth and B. I. Lundqvist, *Phys. Rev. Lett.*, 2004, **92**, 246401.

[341] J. Klimeš, D. R. Bowler and A. Michaelides, *J. Phys. Condens. Matter*, 2009, **22**, 022201.

[342] H. Rydberg, B. I. Lundqvist, D. C. Langreth and M. Dion, *Phys. Rev. B*, 2000, **62**, 6997–7006.

[343] O. A. Von Lilienfeld, I. Tavernelli, U. Rothlisberger and D. Sebastiani, *Phys. Rev. Lett.*, 2004, **93**, 153004.

[344] E. R. Johnson, I. D. Mackie and G. A. DiLabio, *J. Phys. Org. Chem.*, 2009, **22**, 1127–1135.

[345] Y. Sun, Y.-H. Kim, K. Lee and S. Zhang, *J. Chem. Phys.*, 2008, **129**, 154102.

[346] E. Suarez, N. Diaz and D. Suarez, *J. Chem. Theory Comput.*, 2011, **7**, 2638–2653.

[347] J. A. Pople, A. P. Scott, M. W. Wong and L. Radom, *Isr. J. Chem.*, 1993, **33**, 345–350.

[348] R. D. Johnson III, K. K. Irikura, R. N. Kacker and R. Kessel, *J. Chem. Theory Comput.*, 2010, **6**, 2822–2828.

[349] P. Sinha, S. E. Boesch, C. Gu, R. A. Wheeler and A. K. Wilson, *J. Phys. Chem. A*, 2004, **108**, 9213–9217.

[350] M. D. Halls, J. Velkovski and H. B. Schlegel, *Theor. Chem. Acc.*, 2001, **105**, 413–421.

[351] J. M. Bowman, *J. Chem. Phys.*, 1978, **68**, 608–610.

[352] L. S. Norris, M. A. Ratner, A. E. Roitberg and R. B. Gerber, *J. Chem. Phys.*, 1996, **105**, 11261–11267.

[353] A. Willetts, N. C. Handy, W. H. Green and D. Jayatilaka, *J. Phys. Chem.*, 1990, **94**, 5608–5616.

[354] B. Temelso and G. C. Shields, *J. Chem. Theory Comput.*, 2011, **7**, 2804–2817.

[355] B. Temelso, K. A. Archer and G. C. Shields, *J. Phys. Chem. A*, 2011, **115**, 12034–12046.

[356] K. S. Pitzer and W. D. Gwinn, *J. Chem. Phys.*, 1942, **10**, 428–440.

[357] G. Katzer and A. F. Sax, *J. Comp. Chem.*, 2005, **26**, 1438–1451.

[358] P. Y. Ayala and H. B. Schlegel, *J. Chem. Phys.*, 1998, **108**, 2314–2325.

[359] D. G. Truhlar, *J. Comp. Chem.*, 1991, **12**, 266–270.

[360] A. L. L. East and L. Radom, *J. Chem. Phys.*, 1997, **106**, 6655–6674.

[361] W. M. F. Fabian, *Monatsh. Chem.*, 2008, **139**, 309–318.

[362] M. Strekalov, *Chem. Phys.*, 2009, **355**, 62–66.

[363] J. Pfaendtner, X. Yu and L. J. Broadbelt, *Theor. Chem. Acc.*, 2007, **118**, 881.

[364] S. Grimme, *Chem. Eur. J.*, 2012, **18**, 9955–9964.

[365] Y.-P. Li, J. Gomes, S. M. Sharada, A. T. Bell and M. Head-Gordon, *J. Phys. Chem. C*, 2015, **119**, 1840–1850.

[366] R. Sure, J. Antony and S. Grimme, *J. Phys. Chem. B*, 2014, **118**, 3431–3440.

[367] I. Funes-Ardois and R. Paton, *GoodVibes: GoodVibes v1.0.1*, 2016, `http://dx.doi.org/10.5281/zenodo.60811`.

[368] J. Elm and K. V. Mikkelsen, *Chem. Phys. Lett.*, 2014, **615**, 26–29.

[369] T. Kurtén, M. R. Sundberg, H. Vehkamäki, M. Noppel, J. Blomqvist and M. Kulmala, *J. Phys. Chem. A*, 2006, **110**, 7178–7188.

[370] J. Herb, A. B. Nadykto and F. Yu, *Chem. Phys. Lett.*, 2011, **518**, 7–14.

[371] A. B. Nadykto, F. Yu and J. Herb, *Int. J. Mol. Sci.*, 2008, **9**, 2184–2193.

[372] A. B. Nadykto, J. Herb, F. Yu and Y. Xu, *Chem. Phys. Lett.*, 2014, **609**, 42–49.

[373] Y. Xu, A. B. Nadykto, F. Yu, L. Jiang and W. Wang, *J. Mol. Struc.: THEOCHEM*, 2010, **951**, 28–33.

[374] A. S. Hansen, Z. Maroun, K. Mackeprang, B. N. Frandsen and H. G. Kjærgaard, *Phys. Chem. Chem. Phys.*, 2016, **18**, 23831–23839.

[375] A. B. Nadykto, J. Herb, F. Yu, E. S. Nazarenko and Y. Xu, *Chem. Phys. Lett.*, 2015, **624**, 111–118.

[376] A. Nadykto, K. Nazarenko, P. Markov and F. Yu, AIP Conference Proceedings, 2016, p. 090010.

[377] O. Kupiainen, I. K. Ortega, T. Kurtén and H. Vehkamäki, *Atmos. Chem. Phys.*, 2012, **12**, 3591–3599.

[378] H. Henschel, J. C. A. Navarro, T. Yli-Juuti, O. Kupiainen-Määttä, T. Olenius, I. K. Ortega, S. L. Clegg, T. Kurtén, I. Riipinen and H. Vehkamäki, *J. Phys. Chem. A*, 2014, **118**, 2599–2611.

[379] T. Olenius, T. Kurtén, O. Kupiainen-Määttä, H. Henschel, I. K. Ortega and H. Vehkamäki, *Aerosol Sci. Technol.*, 2014, **48**, 593–603.

[380] T. Olenius, R. Halonen, T. Kurtén, H. Henschel, O. Kupiainen-Määttä, I. K. Ortega, C. N. Jen, H. Vehkamäki and I. Riipinen, *J. Geophys. Res. Atmos.*, 2017, **122**, 7103–7118.

[381] M. Pabst, A. Köhn, J. Gauss and J. F. Stanton, *Chem. Phys. Lett.*, 2010, **495**, 135–140.

[382] J. A. Montgomery Jr, M. J. Frisch, J. W. Ochterski and G. A. Petersson, *J. Chem. Phys.*, 1999, **110**, 2822–2827.

[383] N. Chéron, D. Jacquemin and P. Fleurat-Lessard, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7170–7175.

[384] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox, *Gaussian09 Revision D.01*, Gaussian Inc. Wallingford CT 2009.

[385] H.-J. Werner, P. J. Knowles, G. Knizia, F. R. Manby and M. Schütz, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2012, **2**, 242–253.

[386] W. J. Hehre, R. Ditchfield and J. A. Pople, *J. Chem. Phys.*, 1972, **56**, 2257–2261.

[387] M. J. Frisch, J. A. Pople and J. S. Binkley, *J. Chem. Phys.*, 1984, **80**, 3265–3269.

[388] T. H. Dunning, K. A. Peterson and A. K. Wilson, *J. Chem. Phys.*, 2001, **114**, 9244–9253.

[389] R. Krishnan, J. S. Binkley, R. Seeger and J. A. Pople, *J. Chem. Phys.*, 1980, **72**, 650–654.

[390] E. Papajak and D. G. Truhlar, *J. Chem. Theory Comput.*, 2011, **7**, 10–18.

[391] J. Elm, T. Kurtén, M. Bilde and K. V. Mikkelsen, *J. Phys. Chem. A*, 2014, **118**, 7892–7900.

[392] M. Camredon, J. F. Hamilton, M. S. Alam, K. P. Wyche, T. Carr, I. R. White, P. S. Monks, A. R. Rickard and W. J. Bloss, *Atmos. Chem. Phys.*, 2010, **10**, 2893–2917.

[393] F. Yasmeen, R. Vermeylen, R. Szmigielski, Y. Iinuma, O. Böge, H. Herrmann, W. Maenhaut and M. Claeys, *Atmos. Chem. Phys.*, 2010, **10**, 9383–9392.

[394] J. D. Surratt, J. H. Kroll, T. E. Kleindienst, E. O. Edney, M. Claeys, A. Sorooshian, N. L. Ng, J. H. Offenberg, M. Lewandowski, M. Jaoui, R. C. Flagan and J. H. Seinfeld, *Environ. Sci. Technol.*, 2007, **41**, 517–527.

[395] M. Kalberer, M. Sax and V. Samburova, *Environ. Sci. Technol.*, 2006, **40**, 5917–5922.

[396] K. A. Denkenberger, R. C. Moffet, J. C. Holecek, T. P. Rebotier and K. A. Prather, *Environ. Sci. Technol.*, 2007, **41**, 5439–5446.

[397] K. Altieri, S. Seitzinger, A. Carlton, B. Turpin, G. Klein and A. Marshall, *Atmos. Environ.*, 2008, **42**, 1476–1490.

[398] S. Gao, N. L. Ng, M. Keywood, V. Varutbangkul, R. Bahreini, A. Nenes, J. He, K. Y. Yoo, J. L. Beauchamp, R. P. Hodyss, R. C. Flagan and J. H. Seinfeld, *Environ. Sci. Technol.*, 2004, **38**, 6582–6589.

[399] C. Mohr, F. D. Lopez-Hilfiker, T. Yli-Juuti, A. Heitto, A. Lutz, M. Hallquist, E. L. D'Ambro, M. P. Rissanen, L. Hao, S. Schobesberger, M. Kulmala, R. L. Mauldin, U. Makkonen, M. Sipilä, T. Petäjä and J. A. Thornton, *Geophys. Res. Lett.*, 2017, **44**, 2958–2966.

[400] B. Marescau, D. R. Deshmukh, M. Kockx, I. Possemiers, I. A. Qureshi, P. Wiechert and P. P. D. Deyn, *Metabolism*, 1992, **41**, 526–532.

[401] J. E. Bonas, B. D. Cohen and S. Natelson, *Microchem. J.*, 1963, **7**, 63–77.

[402] R. Kumar, V. Choudhary, S. Mishra, I. Varma and B. Mattiason, *Ind. Crops Prod.*, 2002, **16**, 155–172.

[403] J. C. Oxley, J. L. Smith, S. Naik and J. Moran, *J. Eng. Mater.*, 2008, **27**, 17–39.

[404] E. D. Raczynska, M. K. Cyranski, M. Gutowski, J. Rak, J.-F. Gal, P.-C. Maria, M. Darowska and K. Duczmal, *J. Phys. Org. Chem.*, 2003, **16**, 91–106.